P
Agric
R
S

# ROTHAMSTED MEMOIRS

(FROM THE 'PHILOSOPHICAL TRANSACTIONS.')

VOLUME IV.
1914—1934

(Quarto Series)

313829/35
3
27

ROTHAMSTED EXPERIMENTAL STATION, HARPENDEN.

ST. ALBANS:
GIBBS AND BAMFORTH LTD., SPENCER STREET.
1934.

# ROTHAMSTED MEMOIRS

CONTENTS OF VOLUME IV. (Quarto Series).


| | | Published |
|---|---|---|
| 1. | "Some Notes on Soil Protozoa." C. H. MARTIN and K. R. LEWIN. Phil. Trans. B. Vol. CCV., pp. 77-94 ... ... ... ... | 1914 |
| 2. | "On a Form of *Botrytis Cinerea*, with colourless Sclerotia." W. B. BRIERLEY. Phil. Trans. B. Vol. CCX., pp. 83-114. ... ... ... | 1920 |
| 3. | "On the Mathematical Foundations of Theoretical Statistics." R. A. FISHER. Phil. Trans. A, Vol. CCXXII., pp. 309-368. ... ... | 1922 |
| 4. | "A Quantitative Investigation of the Bacterial and Protozoan Population of the Soil, with an Account of the Protozoan Fauna." D. WARD CUTLER, L. M. CRUMP and H. SANDON. Phil. Trans. B. Vol. CCXI., pp. 317-350. ... ... ... ... ... ... ... | 1922 |
| 5. | "The Influence of Rainfall on the Yield of Wheat at Rothamsted." R. A. FISHER. Phil. Trans. B. Vol. CCXIII., pp 89-142. ... ... | 1924 |
| 6. | "The Interaction of Clay with Water and Organic Liquids as Measured by Specific Volume Changes and its Relation to the Phenomena of Crumb Formation in Soils." E. W. RUSSELL. Phil. Trans. A. Vol. CCXXXIII., pp. 361-389. ... ... ... ... ... ... | 1934 |


---

The following paper is not included in the Volume, as it also appeared in Jour. Agric. Sci., Vol. VI., pp. 278-301 (1914). See Rothamsted Memoirs, Vol. VI., Octavo Series :—


"The Soil Solution and the Mineral Constituents of the Soil." A. D. HALL, W. E. BRENCHLEY and L. M. UNDERWOOD. Phil. Trans. B. Vol. CCIV., pp. 179-200. ... ... ... ... ... ... 1914

# III. *Some Notes on Soil Protozoa.*

*By* C. H. Martin, *M.A.*, *and* K. R. Lewin, *B.A.*, *Rothamsted Experimental Station.*

*Communicated by* Horace T. Brown, *F.R.S.*

(Received December 5, 1913,—Read February 5, 1914.)

[Plates 5 and 6.]

Contents.

PAGE

I. Introduction . . . . . 77
II. Methods of Preparation and Culture . . . . . 80
III. Cucumber bed—*Vahlkampfia soli*, n. sp., *Amœba cucumis*, n. sp. . . . . 82
IV. Seedling bed—*Amœba gobanniensis*, n. sp. . . . . 89
V. Conclusions . . . . . 91
VI. Literature List . . . . . 92
VII. Description of Plates . . . . . 93

## Introduction.

It is rather a curious fact that it has been left to agricultural chemists to bring into prominence the important part that free-living Protozoa may play in the soil. At the present time it seems to us that the prevalent idea as regards the distribution of free-living protozoa is that they can exist in the trophic state only in definite accumulations of fresh water (*e.g.*, pools, rivers, lakes) or in the sea. It seems also to be generally held that after the drying up of small pools the cysts of protozoa are transported by currents of air, and that it is to these wind-borne cysts that the prevalence of protozoa in cultures of ordinary soils may be mostly attributed. It is very difficult to trace the historical origin of this view, but it is evident that to the early workers on protozoa the idea of this limitation of their active life to water was unthought of.

It is unnecessary to refer here to the works of very early writers, by whom the spread of disease was sometimes attributed to invisible forms of life present in the air. It would seem, however, from the following quotation that Ehrenberg held a far wider view as regards the distribution of protozoa than that which is fashionable at present. In his great work 'Die Infusionsthierchen als Vollkommene Organismen,' p. 496, he says :—

"Jeder der bekannten lebenden Körper besitzt eine Organisations-Feuchtigkeit. So lange er diese, den ihn bestürmenden physikalischen Naturkräften entgegenkämpfend, in seinen Hauptorganen erhält, so lange ist er lebend ; sobald sie durch

 Published separately, May 30, 1914.

Hitze Frost oder eigene Schwäche verloren geht, oder durch und durch erstarrt, erfolgt der Tod, der auch auf manche andere Weise eintreten kann. Diese Organisations-Feuchtigkeit nehmen Käferlarven im dürrsten Holze, Mottenlarven im dürrsten Pelze, Infusorien und Mooswurzeln, Samen, dergl., im dürrsten Sande aus dem Dunste der Atmosphäre in sich auf, bleiben fleischig und feucht und nässen sogar ihre Umgebung. Lebende Dammerde bleibt feucht."

It is important to note that at the time this work was written the cysts of Protozoa were still unknown, and that therefore, for EHRENBERG, life meant trophic life, and not latent life in the cyst.

By later workers this wide view of the distribution of protozoa seems to have become gradually more limited. For example, DUJARDIN in his account of the protozoa in the 'Histoire Naturelle des Zoophytes,' 1841, p. 169, states:—

"Certains Infusoires vivent, non pas simplement dans les eaux, mais dans les sites habituellement humectés, comme les touffes de mousses, et surtout les couches minces d'oscillaires, sur la terre ou sur les murs humides; pour les trouver, il suffit d'agiter et de presser dans un vase d'eau successivement plusieurs touffes de mousse prise au pied des arbres, dans les lieux frais, ou au bord des ruisseaux; ou bien de placer dans une soucoupe, avec un peu d'eau, la pellicule enlevée à la surface du sol couvert d'oscillaires."

From this quotation it is evident that DUJARDIN regarded the existence of the Protozoa in their trophic state in some saturated soils as certain, but even in the case of saturated soils later writers seem to have believed that the forms met under these conditions were mostly present as cysts. It is to STEIN that we owe the discovery that nearly all free-living protozoa can encyst, and it is probably to him also that we owe the present limitation of our view as to the distribution of protozoa. On p. 20 of his work 'Der Organismus der Flagellaten' (1878) he states:—

"Die encystirten Infusionsthiere erwachen, wenn sie früher oder später unter Wasser gesetzt werden, zu neuem Leben und durchbrechen meist schon nach wenigen Stunden ihre sich erweichende und aufquellende Cyste. Sie können aber auch in trockenem Zustande, gleich Pflanzensamen und Pflanzensporen, durch die Winde auf weite Entfernungen hin fortgeführt und auf den verschiedensten Gegenständen, z. B. auf dem ausgebreiteten Heu der Wiesen, auf den Moosüberzügen alter Gemäuer, zwischen der rissigen Rinde alter Bäume oder im Sande der Dächer abgesetzt werden. Uebergiesst man dann solche Gegenstände mit Wasser, so liefern die daran haftenden Cysten ebenfalls schon nach wenigen Stunden die freie bewegliche Infusorienform."

This view seems to have appealed to BÜTSCHLI since in his work on Protozoa (1 Abtheilung, Sarcodina und Sporozoa, 1880–82) he states on p. 162:—

"So treffen wir Amöben in feuchtem Sand oder Moos von Bäumen, und zwar sowohl am Fusse solcher als in ziemlicher Höhe über dem Erdboden an . . . .

"Dass es sich in diesen Fällen meist um Formen handelt, die durch Winde im encystirten oder zum Theil vielleicht auch nicht encystirten Zustand gewissermaassen verschlagen wurden, dürfte keinem Zweifel unterliegen.

"In dieselbe Kategorie dürfen wir vielleicht auch die von CIENKOWSKY auf Pferdemist beobachtete *Diplophrys stercoreum* und das unter ähnlichen Verhältnissen getroffene *Platoum stercoreum*, sowie den jedenfalls zur gleichen Gattung gehörigen, von GABRIEL in feuchter, mit thierischen Excrementen durchsetzter Erde gefundenen sogen. Troglodytes rechnen, deren nächste Verwandte ja das süsse Wasser bewohnen."

Finally it is to this view of the distribution of protozoa that we probably owe a recent interesting paper by E. M. PUSCHKAREW.* PUSCHKAREW left sterilised plates exposed to the air and examined the cultures that he so obtained. Of the 13 forms he obtained in this way all can be found in soil, but it seems fairly clear that he believed that nearly all these forms in his cultures owed their origin to cysts from dried pools on the banks of the river which had been transported by the wind, *cp.* p. 331.

"Herbst und Winter 1910/11 waren nun zum Teil sehr regnerisch und also ungünstig für meine Arbeit. Nun kam aber das Frühjahr 1911 mit relativ wenig Regen und endlich der Sommer mit seiner grossen Hitze; viele Sümpfe waren ganz ausgetrocknet; das Wasser in Flüssen und Seen stand kolossal niedrig; der meistens mit Algen bewachsene Fluss- und Seeboden war zum Teil von Wasser befreit; die ausgetrockneten Algen waren mit Protozoencysten bedeckt und die Luftströmungen konnten letztere, zusammen mit feinen Schlammteilchen, vom Boden in die Höhe reissen und zu entfernten Orten transportieren."

Of recent years a large amount of our knowledge of that difficult group the Amœbæ has been obtained from animals cultivated from soils (NÄGLER and GLÄSER), but it is interesting to note that little if any attention has been paid to the biology of the forms thus obtained, in their natural state; and we believe that to most biologists the idea that ordinary soils contain a large number of protozoa in an active condition is a new one. The view that the presence of such protozoa in excessive numbers may be the cause of soil sickness was, we believe, first put forward by RUSSELL and HUTCHINSON in a paper "On the Effect of Partial Sterilisation of Soil on the Production of Plant-Food."†

In working with these soil-protozoa, so long as one was confined to culture methods it was obviously impossible to decide how far any given culture of protozoa gave a true picture of the life in a soil, and how far the organisms in the culture were simply derived from cysts. There were, of course, certain indications, as one of us has pointed out in a previous paper, that certain faunas were connected with certain soils,

* "Ueber die Verbreitung der Süsswasserprotozoen durch die Luft," 'Archiv für Protistenkunde,' Band 28.

† 'Journal of Agricultural Science,' vol. 3, Part II (1909).

but even then one was left absolutely in the dark as to what degree of saturation of water was necessary in these soils before the animals found in the cultures could lead an active life in the soil.

By a method which one of us has described in a recent letter to 'Nature,' No. 2266, vol. 91, 1913, and which is given in detail below, it is possible to obtain results which, though minimal, are definite as to the state of the fauna in any given soil at any moment. Up to the present the only soils we have examined thoroughly by this method are (1) a cucumber bed, and (2) a seedling bed obtained from the layer of soil underneath the turf in a neighbouring field and free from any admixture of manure.

In this paper we propose giving some account of the organisms found in an active state in these two soils by this method, together with some further details on the life-cycles of these forms in cultures. In a future paper we hope to return to this question of the existence of active protozoa in the soil in greater detail.

We should like to take this opportunity of thanking Dr. Russell for the great interest which he has taken in this work. We wish also to thank Prof. Minchin for his great kindness in allowing us to write up our results in his department at the Lister Institute, and Miss Rhodes for the trouble she has taken over the drawings.

## Methods of Preparation and Culture.

The only method by which, so far as we are aware, the presence of active protozoa in the soils can be satisfactorily demonstrated is one which one of us has recently described in a letter to 'Nature.' For this purpose a small quantity of the soil to be examined is added to about an equal quantity of a saturated aqueous solution of picric acid; the mixture is then stirred very thoroughly, so that the protozoa which are situated on the liquid films between the soil particles are freed. The mixture is then allowed to stand for some time (12–24 hours), and it will be found that the scum which rises to the surface contains a proportion of the bacteria and protozoa of the soil. Cover-slip preparations can then be made by floating cover-slips upon the surface of the mixture. For the purpose of this work the cover-slips which one of us described in a paper on "A Note on the Protozoa from Sick Soils, with some Account of the Life-Cycle of a Flagellate Monad,"* are very convenient. The cover-slip smears thus obtained can be put in corrosive† or simply washed out in 70-per-cent. alcohol and then stained and mounted in the ordinary way. If the organisms in the soil are scarce it will be found an advantage for purposes of searching to over-stain in eosin.

There can be little doubt that this method is not an ideal one, and that it only gives minimal results, still in suitable soils quite good results can be obtained, for example, in the cucumber soil, the fauna of which is described below, up to 150 Amœbæ have been obtained on a single film. It is to be hoped, however,

* 'Roy. Soc. Proc.,' B, vol. 85.

† *Vide* Mann's 'Physiological Histology,' p. 74.

that in the near future some better method will be found for dealing with this problem, more particularly in poor clay soils, though at present we must confess it is very hard to discover one.

It would we feel be premature at present to attempt a formal list of the culture media on which soil protozoa flourish. In all cases of cultures of soil protozoa, so far as we are aware, up to the present, as VAHLKAMPF clearly insisted in his paper on the biology, etc., of *Amœba limax** the protozoa feed upon the bacteria of the culture, and hence almost any culture media on which soil bacteria flourish will probably support a large number of protozoa.

Therefore in those cases in which the expression "pure animal culture" is used we only wish to indicate that the culture contained only one form of protozoon, though of course it contained large numbers of bacteria. It may of course be possible in the future to obtain cultures of some saprozoic protozoa free from bacteria, and in certain cases we have found indications that certain Amœbæ show a distinct preference for certain culture media, though here, again, this effect may be a secondary one due to the encouragement of a certain type of bacteria.

Up to the present we have mainly used solid media for our cultures, as we find that they are far more convenient for isolating any given form. We used two types of culture media, one an ordinary agar made up of 1000 c.c. meat extract and 15 grm. of agar; but we have found a culture medium of FRIEDBERGER and REITER described in KOLLE and WASSERMANN'S 'Handbuch der pathogenen Mikro-organismen,' vol. 1, gives very good results for most soil protozoa; it consists of a horse-dung agar made up of three lumps of horse dung and 500 c.c. of water, this mixture is boiled for one and a half hours, then filtered through cloth, and finally about 8 grm. of agar is added. In many cases where it is used to get a very strong growth of protozoa it is advisable to add a small amount of water or dilute albumen to the culture plates to about a depth of 2 mm. This addition of water seems to obviate the vacuolated appearance which some workers have noted as characteristic of culture Amœbæ on plates.

The stock cultures are made up by adding a little soil directly to the plates. If these stock cultures are examined from time to time it will be found that in any given culture there is a more or less definite succession of animal forms. By selecting the time and method of culture it will probably be found possible to get pure animal cultures of any of these forms.

In this connection it may be interesting to note the results obtained from two cultures which would seem to indicate that the protozoa in the soil are to a great extent feeders on bacteria, and that the number of bacteria in the mud from ordinary fresh water pools does not seem to offer equal scope for true feeders on bacteria. Two cultures were made from the cucumber soil on manure agar to which, in place of water, Molisch's solution was added. In these cases the algæ got the upper hand

* 'Arch. Protistenk.,' vol. 5.

of the bacteria on the culture, and possibly as a result of this the culture from a protozoon point of view was a very poor one. The counter-experiment was made with some of the cultures of *Amœba proteus*, which were placed with some mud on manure-agar plates, and here the bacteria did not develop to anything like the extent to which they would have done from an ordinary soil culture, and none of the ordinary soil protozoa were found.

### Cucumber Bed.

In July, 1913, through the kindness of a grower in the neighbourhood of Rothamsted we had an opportunity of examining a sick cucumber border. In order to give a fair picture of the very special conditions prevailing in these cucumber beds we have thought it well to give the history of this particular border bed from its origin. It was made up in February, 1912, of one part by volume of light pasture soil, one part of heavy pasture soil, and two parts by volume of horse manure. During the season (April to June 15) the beds were either mulched and given in alternate weeks a small dose of chemical food or were given a dressing of soil, cow-dung, and soot. After June 15 the same treatment was applied with longer intervals. Towards the end of the year two-thirds of the bed was removed and replaced by horse manure. On December 27, 1912, the bed was steam sterilised (20 minutes on a steam grid) and then treated as in the previous year. The water content on the date on which we first examined the bed (August 1, 1913) was in the top layer 62 per cent. by weight, and in the bottom layer 55 per cent. On October 7, 1913, the water content was 57 per cent. in a sample from the bottom. Probably the water content does not vary much in these beds. At first sight this water content may seem high to those who have had no experience of estimations of soil moisture, but the soil at the time of examination was only damp to the touch and there was no sign of excessive water.

Preparations were made of the fresh soil by the picric acid method described above, and the following organisms were discovered in the trophic condition on the films so prepared :—

1. *Euglypha*, ? sp. (Plate 5, fig. 1).
2. *Trinema*, ? sp.
3. *Vahlkampfia soli*, n. sp. (Plate 5, fig. 10).
4. *Amœba cucumis*, n. sp. (Plate 6, fig. 20).
5. Flagellate, ? gen. (Plate 5, fig. 3).
6. *Chilodon* sp. (Plate 5, fig. 2).
7. Ciliate, ? gen.
8. Ciliate, ? gen.

Of these the thecamœbæ were probably present in this soil in considerable numbers, more than is indicated by the number of forms obtained on the films.

This is probably due to the fact that under the most favourable circumstances it is rather difficult to get good preparations of these thecamœbæ from a numerical point of view by the cover-slip method.

The Amœbæ were present on the films in large numbers, *e.g.*, up to 150 good

specimens in an active trophic condition, with large numbers of ingested bacteria, were counted on a single film. These were probably the dominant protozoa in this soil during the time it was under examination.

The flagellates were very rare on these films, and this was rather curious, as in cultures from this soil they were very numerous (Prowazekia and Copromonas, etc.). It would seem that this result can be explained by one of the two following hypotheses:—

(*a*) The films gave a true picture of the life in this soil, and flagellate forms were not present in large numbers in the trophic condition, possibly because the water-content of the soil was not large enough to allow them sufficient liberty for long free paths; or

(*b*) The flagellates were present in an active condition, but the films did not give a true picture, because the long flagella might have become entangled in the soil particles, and thus have prevented the flagellates being carried to the surface-film.

At present we are inclined to accept the former of these views, because in preparations made in the above manner from other soils we have found fair numbers of flagellates in the trophic phase.

The ciliates in these preparations are also very rare, and here it seems probable that their scarcity on the films represents fairly the state of affairs in this soil. It is possible that in absolutely saturated soils ciliates may play an active part as a bacterial check, but it is difficult to believe that they can exercise an important *rôle* in a sick soil-bed like the one under examination. On three occasions ciliates were found in fresh soil-smears: one appeared to be a Chilodon (Plate 5, fig. 2), and the two other forms have not been identified with any certainty.

Of these forms we only propose to deal at length with the two Amœbæ, since not only were they probably the most important checks on the bacteria in this soil at the time of examination, but they also gave the most interesting results on the cultures.

### *Vahlkampfia soli* (n. sp.).

To judge from the evidence of the fresh films, this must have been the dominant form in the cucumber bed during the month of August. Although this animal agrees in many important points with the *Amœba limax* of Vahlkampf's classical paper, yet there are certain small points of difference which seem to justify the formation of a new species for the reception of these forms.

In life *Vahlkampfia soli* is a very active form, showing the characteristic movements of an amœba of the *limax* group. In progression a single, large pseudopodium, composed almost entirely of ectoplasm, is usually thrown out, and the granular entoplasm seems to stream into it. Very often the posterior end shows a characteristic tufted appearance, and it is interesting to note that this appearance is often found in animals on fresh smears (Plate 5. fig. 10), showing that these forms were actively moving through the soil at the time they were fixed by the addition of

picric acid. A contractile vacuole is present, and is usually discharged at the posterior end. The period of active life on these culture-plates at 18° C. seems to be about 3–4 days. In life certain highly refractile granules are fairly frequent in the endoplasm, especially of young forms. In fresh smears prepared from this cucumber soil by the picric acid method, the trophic stages of this animal were very abundant during the month of August. The forms thus found were usually very active, and contained large numbers of ingested bacteria (Plate 5, fig. 10). On some smears indications of division were found (Plate 5, fig. 4), thus clearly showing that *Vahlkampfia soli* flourished in this soil at this date. *Vahlkampfia soli* was found to do fairly well on manure-agar cultures with water, and we also tried plates made up of a straw decoction, which had been recommended by Vahlkampf. In stained forms from young cultures the nucleus is a spherical body, with a large, darkly-staining karyosome of rather elliptical form. The membrane is a fairly definite structure, and ranged upon it there appear to be a number of masses staining with eosin. These forms from young cultures corresponded in every way with the forms found in the fresh films, and it would seem that this appearance of the nucleus is correlated with the intense reproductive activity characteristic of this stage. In older cultures the karyosome is rounder, smaller, and denser; some of the stages of division appear to resemble very closely stages figured by Gläser in the division of *Amœba tachypodia*, and probably some of the stages figured by Aragão for *Amœba diplomitotica*.

It is highly probable, however, that there are a very large number of amœbæ for which this statement will be found true.

As regards the behaviour of the cytoplasm during division, it is perhaps worth noting that the rounding up which Gläser has stated to be always characteristic of dividing amœbæ is not seen in this form, and that the stage of nuclear division does not seem to bear an absolutely fixed relation to the progress of the cytoplasmic division.

The nuclear division of this amœba belongs to that very difficult type, for which Nägler has invented the term of promitosis. In the early stages of division the karyosome swells and becomes very elliptical, and the mass of the chromatin then gets concentrated at either pole of the elongated karyosome. There can be no doubt that the division of the great mass of chromatin contained in the resting nucleus is effected by this simple means. The question of chromosomes in this division is an extraordinarily difficult one, and one cannot help feeling that in divisions of this kind the result arrived at depends much upon the technique adopted, and in many cases it would appear that too much reliance is placed on iron hæmatoxylin preparations alone. In this case we have tried the combined results obtained from iron hæmatoxylin and hæmalum preparations, though we decided, as a matter of caution only, to figure the hæmalum preparations. At a slightly later stage it would appear that the median portion of the karyosome becomes broken on

one side into a mass of irregular granules (Plate 5, fig. 11). This stage seems to correspond in some respects with that figured by Gläser for *Amœba tachypodia*, but in this case he is inclined to refer the origin of these granules to extra-karyosomic chromatin. The dumb-bell-shaped karyosome is now broken through (Plate 5, figs. 12 and 13), and the two polar masses gradually pass to the opposite extremities of the dividing animal (Plate 5, fig. 14). The connecting bar, which seems to consist of a mixture of the chromatin granules mentioned above and an achromatic mass, becomes gradually thinner at its middle, and, as by this process chromatin granules are brought much closer together, it becomes at the same time much more conspicuous (Plate 5, fig. 15). During this stage, for some unknown reason, the chromatin becomes massed in three volumes, viz. the two polar masses, and the median portion of the connecting bar, so that the characteristic appearance shown in Plate 5, fig. 15, is originated by two polar masses, an intervening clear space, and a darkly-stained central bar. The daughter nuclei now become separated through the snapping of this connecting bar, and in the reconstruction of the nuclei it would appear that the wedge-shaped mass derived from the connecting bar loses its chromatin in the form of irregular granules, which fuse with the karyosome (Plate 5, fig. 16). It will be obvious from this account that, though we differ with Gläser over some slight points of detail, yet in this form also we can find no trace of the existence of a centriole.

In life the cysts of these amœbæ are very characteristic objects on the culture-plates. They have an outer gelatinous coat, to which foreign objects become attached, and an inner resistant wall of a yellowish colour. The cysts are frequently grouped together in a culture, and this feature appears to be somewhat important, as upon it some of the earlier workers, *e.g.*, Zopf, place similar forms in the group of Mycetozoa. In his account of *Copromyxa protea* in 'Die Pilzthiere oder Schleimpilze,' Zopf states on p. 133 :—

"Da auf trocknen Mistculturen die Amœben gewöhnlich nicht alle neben, sondern zum Theil über einander kriechen, so entstehen die eingangs erwähnten Sporenhäufchen (Sori) (fig. 31, I, II)."

The mature cysts are extraordinarily difficult to stain. It seems, however, from early stages of encystation, such as is shown in Plate 6, fig. 19, that there is a single central nucleus, and we have no reason to suppose that the cyst is anything but a protective cyst. Under ordinary cultural conditions this form emerges from the cysts as a typical *limax*-amœba ; in the process of excystation the cyst swells considerably and the amœba can be seen moving round actively in the cyst for some time before it emerges through a small opening in the cyst wall.

*The Flagellate Form.*—In 1910, Wasielewski and Hirschfeld pointed out the appearance of certain conditions of a flagellate stage of *limax*-amœbæ, and more recently Whitmore and Wherry have described such stages. In the present case

it was found that, if a plate of manure-agar were heavily infected with the cysts and covered to a depth of about 2 mm. with tap-water containing 0·25 per cent. NaCl and 0·05 per cent. $MgSO_4$, after about 16–20 hours' incubation at 23° C., flagellate forms of this amœba appear sporadically. On three occasions an example of the flagellate was continuously watched under a 2·5 mm. water-immersion objective, and was seen to become amœboid, lose its flagella by absorption, and turn into a typical specimen of the amœba. Permanent preparations of this transition have also been obtained (Plate 6, fig. 18). The flagellate stage (Plate 5, fig. 17) is generally ovoid, with an oblique anterior end, but some specimens are markedly piriform, the pointed end being anterior. The two flagella arise anteriorly and are each about equal in length to the body. The nucleus is situated near the insertion of the flagella, and in stained preparations a fibril connecting the nuclear membrane with the common root of the flagella can be detected.

A contractile vacuole is present, occurring laterally in the posterior third of the body. It arises as a group of small vacuoles, which coalesce to one large one that bulges out the body-wall and eventually bursts like a bubble. Its period is two to three minutes.

The cytoplasm is fairly homogeneous, and at the oblique anterior end there seems to be a slight differentiation recalling ectoplasm. The flagellate is flexible, but shows no tendency to put out pseudopodia.

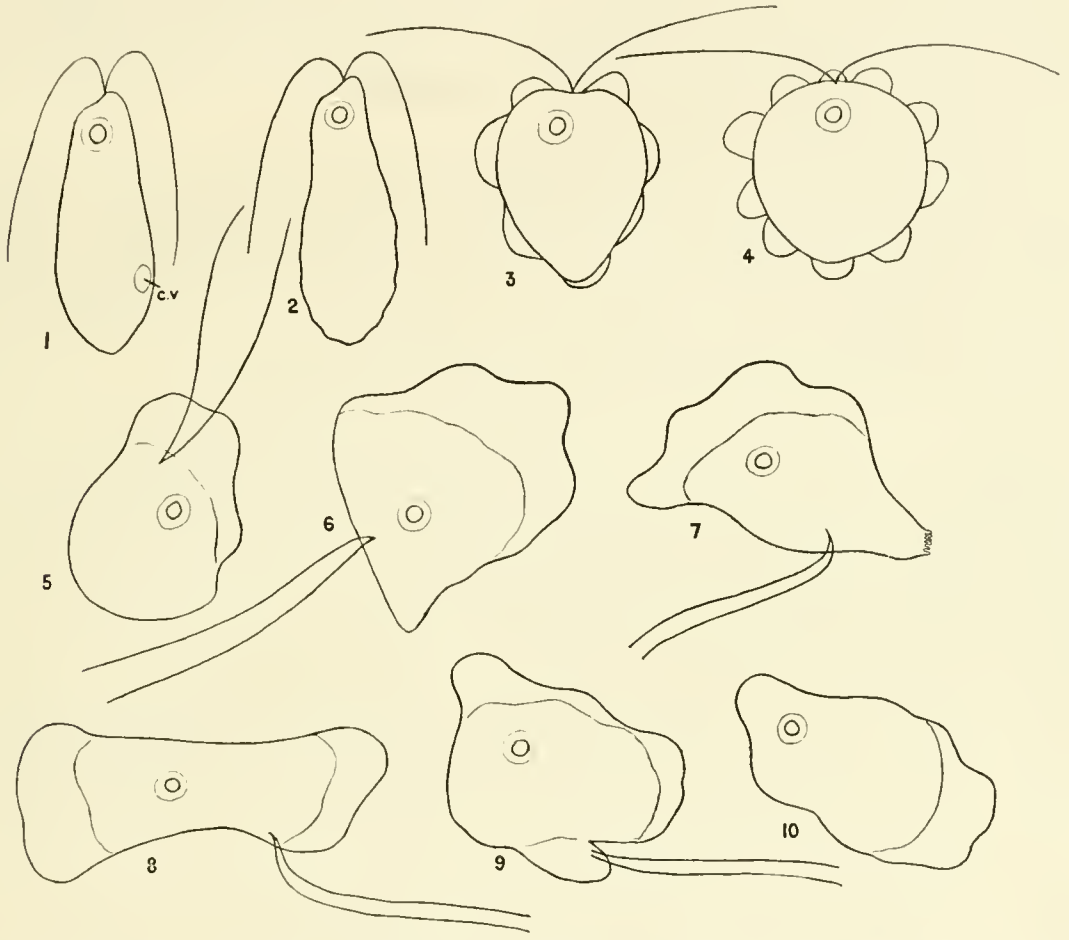
*Movements.*—These are quite rapid, translocation being effected by a steady shouldering movement, during which the animal rotates on its long axis. The flagella are hardly to be made out unless the flagellate be momentarily at rest. They appear to whip right back to the sides of the body.

*Feeding.*—Bacteria are ingested at the oblique anterior surface, where a sort of ectoplasmic differentiation is seen. They pass by the nucleus and are finally digested in the median and posterior parts of the body.

*Transition to Amœba.*—The flagellate remains in one place, and the outlines of its body appear slightly wavy. Then it becomes roughly spherical and protrudes pseudopodia in all directions. Almost at once the spherical form is lost and a typical amœboid appearance is attained. The flagella lose their anterior position and become lateral, or often terminal, but for about seven minutes emerge always in the neighbourhood of the nucleus, to which they are attached. After this time the connection presumably breaks, and the position of the nucleus is no index to that of the flagella, which continually tend towards the hind end at any moment. Shortly after attaining independence, the flagella are active only in a minor degree, and soon become motionless. About 15 minutes (hanging drop preparation; less under cover slip) after the first signs of change, the flagella are absorbed by a pseudopodium which pushes out close to their origin, thrusts them aside, and flows along them and engulfs them.

This process may go some way and then suffer reversal. On one occasion a

flagellate was seen to round up, protrude pseudopodia, and take on typical amœboid form, and after $2\frac{1}{2}$ minutes gradually to recover the definite flagellate shape. After this it was watched continuously for 25 minutes, and then showed no signs of transition. In this instance of the flagellate's attaining the amœboid form, the flagella did not lose their connection with the nucleus.




TEXT-FIG. 1.—Transformation of Flagellated into Amœboid Stage of *V. soli*.
1, Flagellate stage; 2, start of transformation; 3, 1 minute after start; 4, 2 minutes; 5, 3 minutes; 6, 7 minutes; 7, 9 minutes; 8, 12 minutes; 9, 15 minutes; 10, 17 minutes.

The significance of these flagellate stages is at present unknown, but whether their appearance forms grounds for removing the *limax*-amœbæ from the group of true Amœbæ and placing them amongst the Proteomyxa is a question that future work must decide.

### *Amœba cucumis* (n. sp.).

*Amœba cucumis* was also met with in the cucumber soil, and it appeared in the cultures of manure-agar plates associated with *Vahlkampfia soli*. It could be readily distinguished from the Vahlkampfia form in life since it is a very sluggish form with very characteristic pseudopodia. Dividing forms were most frequently found on

cultures three to four days old, and cysts were not frequent until the eighth or ninth day. Active forms associated with cysts have, in fact, been met with in cultures over a month old, though these forms were then of rather small size. In young cultures the typical appearance seems to be an almost spherical form with rather fine radiating pseudopodia composed almost entirely of ectoplasm. Frequently, however, forms are met with in which a kind of wave of ectoplasm is thrown out. As far as can be seen there is no definite contractile vacuole. This form is not nearly such a *typical* devourer of bacteria as *Vahlkampfia soli*, since it not infrequently ingests small flagellates, cysts, etc.

The cytoplasm has a tendency to take up stains rather deeply. The nucleus is almost central, it consists in young cultures of a large, darkly staining karyosome separated by a considerable space from a distinct membrane, and in this space small particles, composed apparently of chromatin, lie, their presence being more noticeable in forms from older cultures. The division of this form seems to show many points of similarity to that described by Gläser for *Amœba lamellipodia*.

In the early stages of division the karyosome becomes much swollen and the cytoplasm of the body becomes very definitely rounded up (Plate 6, fig. 22). The karyosome then becomes broken up into a number of small chromatin masses, and these become arranged in a plate which may extend over half of the diameter of the animal (Plate 6, fig. 23). The chromatin-granules are at first present in a number of rows, but later they become arranged in a double row, thus giving rise to two distinct equatorial plates (Plate 6, fig. 24). These two plates, which lie in a differentiated tract in which we can find no trace of a definite fibril structure, become gradually shifted apart and at the same time more concentrated (Plate 6, fig. 26). At about this stage the first signs of division are seen in the cytoplasmic body. The chromatin-granules now become lumped into masses which coalesce to form a new karyosome. An odd feature of this division is the prominence of a thin thread connecting the daughter-individuals (Plate 6, fig. 28), which seems absolutely identical with that figured by Gläser for *Amœba lamellipodia*. In many cases encystation seemed to be preceded by an association of two individuals correlated with some rather complicated nuclear changes; to these, however, we hope to return on a future occasion. The ripe cyst in life is a thin-walled, colourless structure (Plate 6, fig. 29). The nucleus lies rather eccentrically, and the cytoplasm is at certain stages crowded with darkly staining masses which appear to be extruded by chromatin granules.

In the process of excystation it appears that the cyst-wall is dissolved. For example, an excysting form was picked up at 3.30; at this time the cyst-wall was quite definite and the nucleus rather elongate. Vacuoles now made their appearance at the periphery of the cytoplasm and under the cyst-wall, and by 3.58 the wall was completely dissolved and the typical membranous pseudopodia had made their appearance.

### Seedling Bed.

In March of this year one of us had an opportunity of examining, through the kindness of Mr. Pitt of Abergavenny, the soil of a frame bed in which seedling cauliflowers were being grown. The bed was made this year by taking the first spit of soil after the turf had been removed from a neighbouring field and mixing it with sand and leaf mould; no manure was added, and the frame was not heated, so that the conditions were as different from those in the preceding soil as could be imagined.

The active fauna found by means of the picric acid method was not as rich in the number of individuals but far richer in the number of species than the cucumber soil. This case seems to present an interesting analogy from a faunistic point of view to results obtained on the grass plots at Rothamsted, where, as is well known, the untreated plot still gives a large number of species, whereas in some of the plots which have received heavy manure of one kind for many years the number of species has been cut down to four or five.

The same phenomenon seems to be shown in rich infusions, in which as a rule one or other protozoon gets an upper hand, whereas in ordinary fresh-water pools the fauna, though far richer from the point of view of the number of species to be met with, is far poorer in the actual numbers of individuals. In fresh preparations of this soil, of which only a few were carefully examined, the following were some of the protozoa met with in a trophic condition :—

1. *Euglypha*, ? sp. (Plate 5, fig. 5).
2. *Chlamydophrys*, ? sp. (Plate 5, fig. 6).
3. *Amœba gobanniensis*, n. sp. (Plate 5, fig. 7).
4. *Amœba*, ? sp. (Plate 5, fig. 8).
5. *Amœba*, ? sp. (Plate 5, fig. 9).
6. *Amœba*, ? sp.
7. Flagellate amœba.
8. *Bodo caudatus* (Plate 6, figs. 35, 36, and 37).

On cultures from this soil *Amœba diploidea*, a monad and *Prowazekia* sp. ? and a small ciliate were also met with. It is probable that a more careful search would have revealed *Amœba diploidea* in a trophic condition on fresh films of this soil.

In this paper we only propose to note some of the forms met with in this soil with the exception of an amœba, *Amœba gobanniensis* n. sp., which seems of especial interest since it is evidently so closely allied to the *Amœba cucumis* found in the cucumber soil.

The *Chlamydophrys* found in this soil seems to be rather an interesting form as it differs in respect of its size and behaviour of the pseudopodia from the *Chlamydophrys* which one of us cultured out of a soil two years ago and has kept under culture ever since in the hope of being able to confirm Schaudinn's account of the life-cycle of this form. Further details of this form will probably be published in a future paper.

An amœboid form which was rather interesting from the readiness with which it assumed flagellate condition was met with in this soil. It seemed to divide in the amœboid condition.

A relatively large amœba was also met with in fresh films and in cultures from this soil.

A *Bodo*, which we are inclined to identify with the *Bodo caudatus* of DUJARDIN, was met with both on the fresh smears and in the cultures. As will be remembered, this species is characterised by a peculiar method of multiple division, and it is, as far as we are aware, the only flagellate of a *Bodo* type for which this method of division has been described (Plate 6, figs. 35, 36, and 37).

From the point of view of nomenclature it is interesting to observe that this *Bodo* has a typical kinetonucleus, and this would seem to emphasise a point already brought out by ALEXEIEFF as to the very unsatisfactory state of the classification of the genera *Bodo* and *Prowazekia*. Some stages of the division of this form are shown.

The behaviour of the nucleus seems rather complicated and will be dealt with more fully by one of us in a future paper in the 'Zoologischer Anzeiger.'

*Amœba gobanniensis*, n. sp.

This amœba was met with on the fresh films of the seedling soil and on cultures made from it on manure-agar and water.

*Amœba gobanniensis* is again extraordinarily like the *Amœba lamellipodia* described by GLÄSER (16) and the previously described *Amœba cucumis*, but here again there are certain distinctive features which seem to show the necessity for the formation of a new species, in order to avoid the danger of including two different forms under the same specific name.

In life this amœba is a very sluggish form, and is readily recognised by the extraordinary development of its ectoplasm. The characteristic prickle pseudopodia of *Amœba cucumis* (*cf.* Plate 6, fig. 21) are absent in this form, in which the resting form is characterised by a large plate-like ectoplasmic pseudopodium surrounding the animal. In movement a long axial pseudopodium from the entoplasmic core seems to be thrown out by the animal.

The nucleus in the stained form consists of rather a small karyosome, between which and the membrane lie a number of irregular granules.

The behaviour of this animal in division recalls in many points the division of *Amœba cucumis*, particularly in the last stage of division, in which the dividing products remain for some time connected by a narrow thread of protoplasm (Plate 6, fig. 34). In examining a large number of division stages of both these forms, there seem to be certain constant points of difference. Firstly, the dividing *Amœba gobanniensis* never becomes so spherical as the *Amœba cucumis*, and it is always

characterised by the behaviour of the ectoplasm. Secondly, the band of chromatin granules in the early stage of division is not nearly so broad as in the corresponding stage of *Amœba cucumis*, and in the later stages the spindle seems always slightly smaller. Thirdly, there is a more definite indication of fibrous structure in the spindle of *Amœba gobanniensis* (Plate 6, fig. 32).

The stages leading up to encystation seem to show similar nuclear complications to those that have been noticed in *Amœba cucumis*, and the cyst seems to be very similar.

## Conclusions.

It seems generally agreed that further examination of the Amœbæ will necessitate the splitting up of the genus *Amœba* into a number of genera. The first and much needed step in this reform has already been taken by Chatton (13), by the formation of the genus *Vahlkampfia* for the group of *limax* amœbæ. Whether it would not be better to put this genus and all the other amœbæ which show a flagellate stage in their life-cycle apart from the gamete stage into the group Proteomyxa seems to us an open question. There can be no doubt that in this genus *Vahlkampfia* a number of quite definite species are included which can probably be best separated by minute differences in the behaviour of the nucleus during division. It will probably be found necessary in the same way to form another genus for the *lamellipodia* group of amœbæ, which would again have to be broken up into a number of species in a similar manner.

The main purpose of this introductory paper has not, however, been the study of these amœbæ from a specific point of view, so much as the proof which we hope to have brought of the existence of a relatively frequent trophic Protozoan fauna in certain soils and the rough indication of some possible methods of dealing with this fauna. How far this fauna under certain conditions exercises a deleterious influence on plant growth is rather a question for the agriculturist than the zoologist.

The startling success in the Lee Valley of the treatment of sick soils by partial sterilisation, introduced by Russell, would seem to present a very strong argument in favour of the view that these protozoa do exercise an important influence on plant growth in these soils. We have, by means of the method described above, been able to establish the occurrence of a trophic Protozoan fauna in certain field soils that we have examined, and to this question we hope to return in a future paper.

## LITERATURE.


1. Alexeieff, "La Division nucléaire et l'Enkystement chez quelques Amibes, I–III," 'Compt. Rend. Soc. Biol.,' vol. 70.
2. *Idem*, "Sur le Stade Flagellé dans l'Évolution des Amibes limax," 'Compt. Rend. Soc. Biol.,' vol. 72 (1912).
3. *Idem*, "Sur les Caractères Cytologiques et la Systématique des Amibes du groupe Limax (*Nægleria* nov. gen. et *Hartmannia* nov. gen.) et des Amibes Parasites des Vertébrés (*Proctamœba* nov. gen.)," 'Bull. Soc. Zool. France,' vol. 37 (1912).
4. *Idem*, "Quelques Remarques Complémentaires sur la Systématique des Amibes du groupe Limax," 'Bull. Soc. Zool. France,' vol. 37 (1912).
5. *Idem*, "Systématisation de la Mitose dite 'primitive,'" 'Arch. Protistenk.,' vol. 29 (1913).
6. Aragão, "*Amœba diplomitotica*," 'Mém. do Inst. Oswaldo Cruz,' vol. 1 (1909).
7. Beihla, 'Die Amöben, insbesondere vom parasitären und kulturellen Standpunkte.' Berlin, 1898.
8. Beyerinck, M. W., "Kulturversuche mit Amœben auf festen Substraten," 'Centralb. f. Bakt. u. Parasit., Jena,' vol. 19 (1896), and vol. 21 (1898).
9. Bütschli, "Protozoa," 'Bronn's Thierreich.'
10. Brodsky, "Division and Encystment of *Amœba hyalina* (Davy)," 'Biol. Zeitsch.,' vol. 1 (1910).
11. Chatton, "Sur quelques genres d'Amibes libres et parasites," 'Bull. Soc. Zool. France, vol. 37 (1912).
12. *Idem*, "La Structure du Noyau et la Mitose chez les Amœbiens," 'Archives de Zool. Expér.,' vol. 5 (1910)
13. Chatton and Lalung-Bonnaire, "Une *Amibe Limax* (*Vahlkampfia*, N. G.) dans l'intestin humain," 'Bull. Soc. Path. Exot.,' vol. 5.
14. Dangeard, "Études sur le développement des Organismes inférieurs," 'Le Botaniste,' vol. 11 (1910).
15. Frosch, P., "Zur Frage der Reinzuchtung der Amœben," 'Centralb. f. Bakt. u. Parasit.,' vol. 21.
16. Gläser, "Über die Teilung einiger Amöben," 'Arch. Protistenk.,' vol. 25 (1912).
17. *Idem*, "Kernteilung Encystierung und Reifung von *Amœba mira*," 'Arch. Protistenk.,' vol. 27 (1912).
18. Hartmann and Chagas, "Über die Kernteilung von *Amœba hyalina* (Davy)," 'Mem. Inst. Oswaldo Cruz.,' vol. 2 (1910).
19. Hartmann, M., and Nägler, K., "Copulation bei *Amœba diploidea*, etc.," 'Sitzb. d. Ges. d. Naturf. Freunde.' Berlin, 1908.

20. Nägler, "Entwicklungsgeschichtliche Studien über Amœben," 'Arch. Protistenk.,' vol. 15 (1909).
21. *Idem*, "Studien über Protozoen aus einem Almtumpel—*Amœba Hartmanni*," 'Arch. Protistenk.,' vol. 22 (1911).
22. *Idem*, "Die Kern- und Centriolteilung bei Amöben," 'Arch. Protistenk.,' vol. 26 (1912).
23. Puschkarew, "Über die Verbreitung der Süsswasser-protozoen durch die Luft," 'Arch. Protistenk.,' vol. 28 (1913).
24. Vahlkampf, E., "Biologie und Entwicklungsgechichte von *Amœba limax*," 'Arch. Protistenk.,' vol. 5, p. 167 (1905).
25. Walker, "Die Technik der Amœbenzuchtung," 'Centralb. für Bakter.,' vol. 50 (1911).
26. Wasielewski and Hirschfeld, "Untersuchungen ü. Kulturamöben," 'Abh. Heidelberg. Akad. Wiss. Math.-naturw. Kl.' (1910).
27. Wherry, "Studies on the Biology of an Amœba of the Limax group," 'Arch. Protistenk.,' vol. 31 (1913).
28. Whitmore, "Studien über Kulturamöben aus Manila," 'Arch. Protistenk.,' vol. 23 (1911).
29. Zopf, "Die Pilztiere oder Schleimpilze," 'Encyclopädie der Naturwiss.,' Breslau, 1885.

## DESCRIPTION OF PLATES.

### Plate 5.

Fig. 1.—*Euglypha* sp. from fresh preparation of cucumber bed.
,, 2.—*Chilodon* sp. from fresh preparation of cucumber bed.
,, 3.—Flagellate from fresh preparation of cucumber bed.
,, 4.—Dividing *Vahlkampfia soli* from fresh preparation of cucumber bed.
,, 5.—*Euglypha* sp. from fresh preparation of seedling bed.
,, 6.—*Chlamydophrys* sp. from fresh preparation of seedling bed.
,, 7.—*Amœba yobanniensis* from fresh preparation of seedling bed.
,, 8.—*Amœba* sp. from fresh preparation of seedling bed.
,, 9.—*Amœba* sp. from fresh preparation of seedling bed.

### *Vahlkampfia soli* (n. sp.).

(All the drawings of this form were made under Zeiss 2 mm. apoch. + 18 comp. oc.)

Fig. 10.—*Vahlkampfia soli* from fresh preparation of cucumber bed.
Figs. 11–16.—Stages in division of *Vahlkampfia soli*.
Fig. 17.—Flagellated stage of *Vahlkampfia soli*.

## Plate 6.

*Vahlkampfia soli* (n. sp.).

Fig. 18.—Transition to amœboid stage (whole length of flagella not shown).
„ 19.—Cyst of *Vahlkampfia soli*.

*Amœba cucumis* (n. sp.).

(Zeiss 2 mm. apoch. + 12 comp. oc.)

Fig. 20.—*Amœba cucumis* from fresh preparation of cucumber bed.
„ 21.—*Amœba cucumis* from young culture.
Figs. 22–28.—Stages in division of *Amœba cucumis*.
Fig. 29.—Cyst of *Amœba cucumis*.

*Amœba gobanniensis* (n. sp.).

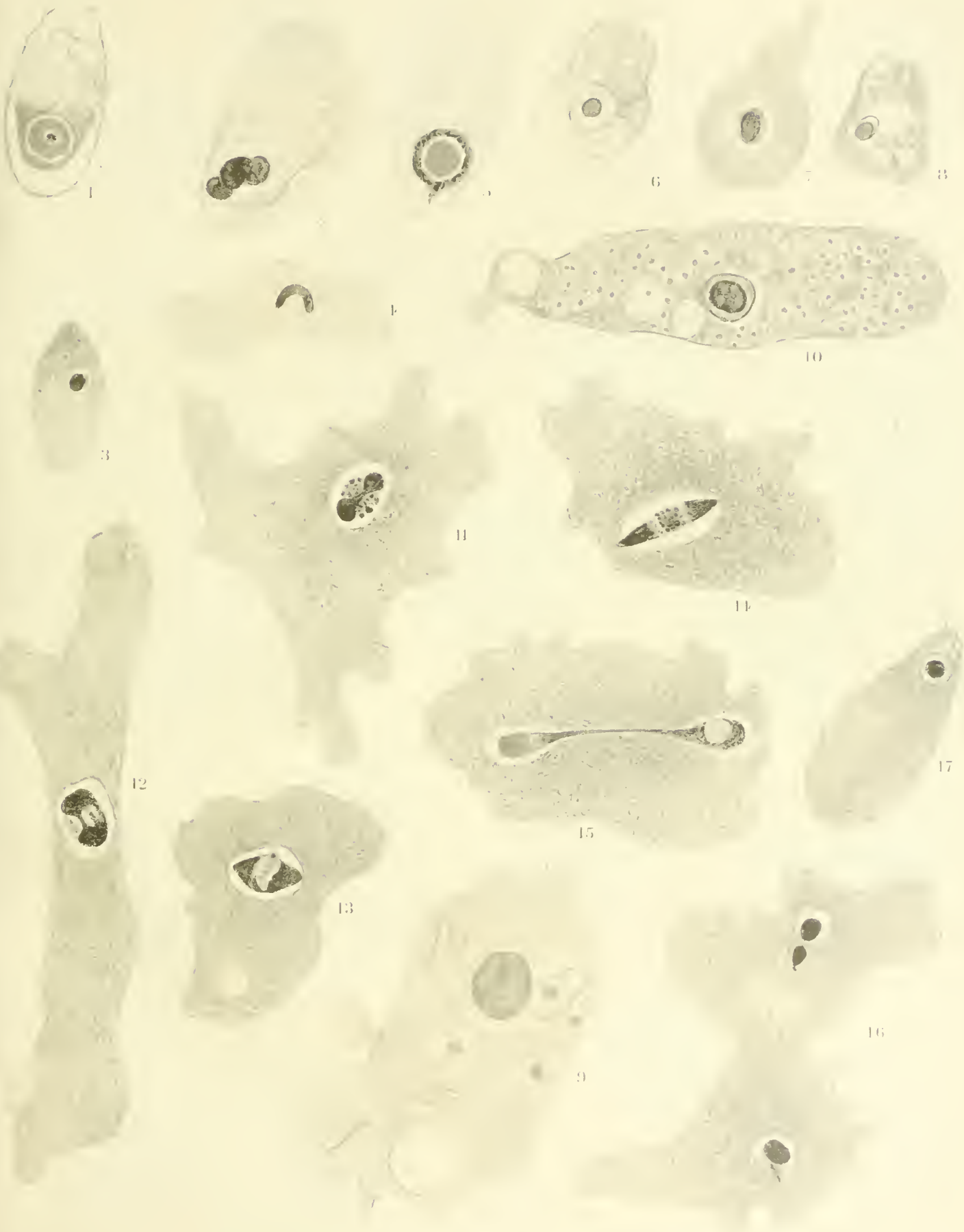
(Zeiss 2 mm. apoch. + 12 comp. oc.)

Fig. 30.—*Amœba gobanniensis* from culture.
Figs. 31–34.—Stages in division of *Amœba gobanniensis*.

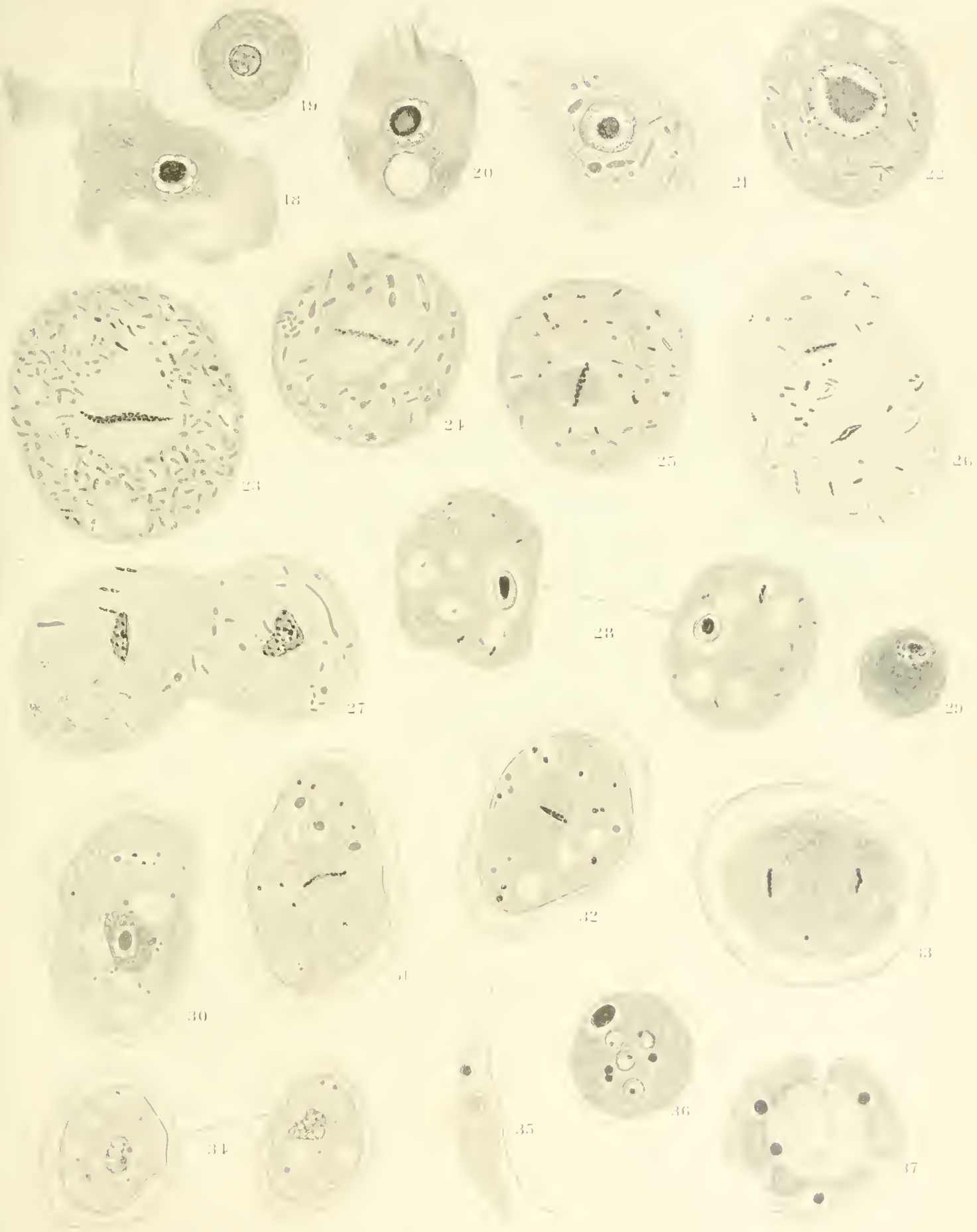
*Bodo caudatus*.

(Zeiss 1·5 mm. apoch. + 18 comp. oc.)

Fig. 35.—*Bodo caudatus*.
„ 36.—Multiple division of *Bodo caudatus*.
„ 37.—Multiple division of *Bodo caudatus*, later stage.




M.K. del


19
18
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

# III.—*On a Form of* Botrytis cinerea, *with Colourless Sclerotia.*

*By* William B. Brierley (*Rothamsted*).

*Communicated by* Dr. E. J. Russell, *F.R.S.*

[Plate 5.]

(Received January 31, 1920,—Read March 18, 1920.)

Contents.

PAGE

1. Introduction . . . . . 83
2. Life cycle of *Botrytis cinerea* . . . . . 84
3. Origin of colourless form: General . . . . . 86
4. Comparison of parental strain and its colourless derivative . . . . . 88
   A. Physiological activities . . . . . 89
      (*a*) Germination of spores . . . . . 89
      (*b*) Growth on nutrient media . . . . . 90
      (*c*) Pathogenicity . . . . . 91
   B. Morphological characters . . . . . 92
      (*d*) Dimensions of spores . . . . . 92
      (*e*) Haptera . . . . . 92
      (*f*) Sclerotia . . . . . 93
5. Origin of colourless form: Special . . . . . 95
   C. Spore formation . . . . . 96
   D. Sclerotial formation . . . . . 99
6. Nature of loss of colour . . . . . 100
7. Significance of colourless form . . . . . 103
   A. Discussion of recorded mutations . . . . . 103
   B. Alternative interpretations . . . . . 107
8. Literature referred to . . . . . 112

## 1. Introduction.

*Botrytis cinerea* is perhaps the commonest and best known fungus, and has been a centre of mycological research since the time of de Bary. Few, if any other fungi, have been studied so thoroughly by so many able investigators, or are the subject of so extensive a literature; and one need only draw attention to the researches of de Bary (6), Pirotta (61), Marshall Ward (80), Kissling (45), Nordhausen (58), R. E. Smith (70), Farneti (27), Beauverie and Guilliermond (9, 10), Istvanffi (33), Reidemeister (65), and the more recent researches from the laboratory of Prof. V. H. Blackman (13, 16, 17, 84). There is a body of experimentally ascertained and exact knowledge concerning the bionomics of this fungus, which can be exceeded by that of few other micro-organisms.

For the particular purposes in view in my investigations, it has been imperative

that all experimental work be carried out with pedigree cultures. It was very early discovered that the ordinary "pure-culture" of mycological and bacteriological laboratories may be and very often is a poly-genotypic population, and that very considerable genotypic differences may usually be found in apparently homogeneous populations exhibiting only one single type, around which the individuals fluctuate. In order to eliminate this vitiating factor, single-spore cultures were prepared either by Burri's (19) Indian-ink method, plate isolation, or, more usually, by a combination of the plate and dilution-drop method. These cultures served as the initial source of experimental material. In all, over seventy such pedigree stocks have been prepared from original sub-strata, representing about sixty different species of host plant, derived from all parts of the country. From these stocks some fifteen thousand cultures have been made. During the course of the investigation, many of these have been subjected to the most diverse environmental conditions, and all have been maintained under the closest scrutiny. With the exception of the single culture to be described, no change that could be interpreted as a permanent heritable alteration or mutation has been observed.

## 2. Life Cycle of *Botrytis cinerea*.

In nature, the fungus usually appears as a delicate smoke-grey velvety pile, covering diseased plant tissues or organic *debris*, or, on more resistant sub-strata, in the form of isolated grey pustules. This superficial growth consists of the branched conidiophores bearing spores. On germination, which under favourable conditions occurs in a few hours, each of the latter gives rise to a mycelium, which in two or three days produces a further crop of conidiophores and spores. When old, the hyphæ and spores may produce microconidia, which germinate directly giving rise to mycelium. If at any stage of development conditions unfavourable to sporogeny intervene, the mycelium produces sclerotia, and these, under more favourable conditions, give rise to tufts of conidiophores bearing spores. Under certain adverse conditions, the hyphæ may produce chlamydospores, or themselves break up into oïdia, and these, on germination, give rise to normal mycelium.

It was the opinion of de Bary (5, 6), Viala (79), Zopf (89), and other early investigators, an opinion based purely on superficial resemblances and the frequent contiguity of growth of the organisms, that *Botrytis cinerea* is only the conidial phase of a Discomycetous fungus, which was referred to either *Sclerotinia Fuckeliana* or *S. libertiana*. In the absence of any proof of this connection, and despite their own negative experience, this opinion was accepted by Marshall Ward (80), Kissling (45), and others, and so became an integral part of mycological literature. In 1905, Istvanffi (32) published a voluminous memoir, purporting to bring forward proof of the genetic relationship of *Botrytis cinerea* and *Sclerotinia Fuckeliana*. In spite of the prior and much more incisive work of R. E. Smith (70), this evidence has been generally accepted. It is, however, of

doubtful value, for a critical examination of the investigation makes clear that Istvanffi has confused two distinct fungi, which possess certain superficial resemblances, often grow together on the same host, and are not easily separable in culture.

Recently, Seaver and Horne (68) claim to have established the relationship of an unnamed species of *Botrytis* with a new species of *Sclerotinia*, which they have termed *S. Geranii*. The evidence is very brief, and awaits confirmation, and, in view of the fact that many "pure-cultures" of fungi are undoubtedly mixed populations, that *Botrytis cinerea* and various species of *Sclerotinia* and other genera harmonise perfectly in their growth when developing intermixed, both on natural hosts and on many artificial media, and that, unless specially searched for, the very minute microconidia of both forms are easily overlooked, and may act as a contaminating factor, one may perhaps hesitate to accept unreservedly this evidence. Furthermore, in my own study, some fifteen thousand cultures of various strains of *Botrytis cinerea* and nearly related "species" have been closely observed under the most varied environmental conditions, and much experimental work has been specially directed towards the elucidation of the genetic relationships of the fungus. I have, however, found no evidence indicating that *Botrytis cinerea* is in any way a developmental phase within the life-cycle of *Sclerotinia Fuckeliana* or genetically related to any other species of this genus. Repeatedly, strains of *Botrytis cinerea* and various species of *Sclerotinia* have been grown side by side by inoculating alternate quadrants of a plate culture, and, under all conditions, have remained separate. My experience merely confirms that of Lind (47), R. E. Smith (70), Peltier (59), Pethybridge (60), and others, who have paid special attention to this aspect of the problem.

*Botrytis cinerea* as a discrete entity is an asexual fungus, and the critical importance of a *Sclerotinial* relationship lies in the possibility which this introduces of a sexual process. If the organism be sexual, any single individual may possibly be heterozygous, and there is then no inherent improbability that segregation may occur, resulting in the appearance of apparently new forms which might mistakenly be interpreted as mutants.

In many *Mucorineæ*, *Chytridineæ*, *Saccharomyces* and possibly other *Ascomycetes* the sexual process is allogamous and factorial segregation is not impossible. In the great majority of fungi, however, in which sexuality occurs, and here is included so far as is known the entire group *Discomycetineæ*, which contains the genus *Sclerotinia*, the process is autogamous. Now it is well known, and has recently been mathematically demonstrated by Jennings (35), that even in an originally heterozygous organism, self-fertilisation if continued generation after generation leads rapidly to a condition in which the offspring are homozygous. In such case Johannsen (36, 37) has shown that the mere isolation of the progeny from a single reproductive member produces a pure line.

Even, therefore, in the remote contingency that *Botrytis cinerea* is merely a developmental phase of *Sclerotinia Fuckeliana* and that this possesses a sexual process—which has yet to be demonstrated for any species of this genus—it would appear that the complications resulting from heterozygosis are absent from our problem. For critical purposes *Botrytis cinerea* is, on all evidential criteria, an asexual homozygotic organism in which the isolation of a single spore strain necessarily implies the isolation of a "pure-line." A genotypic change in a pure line is a mutation.

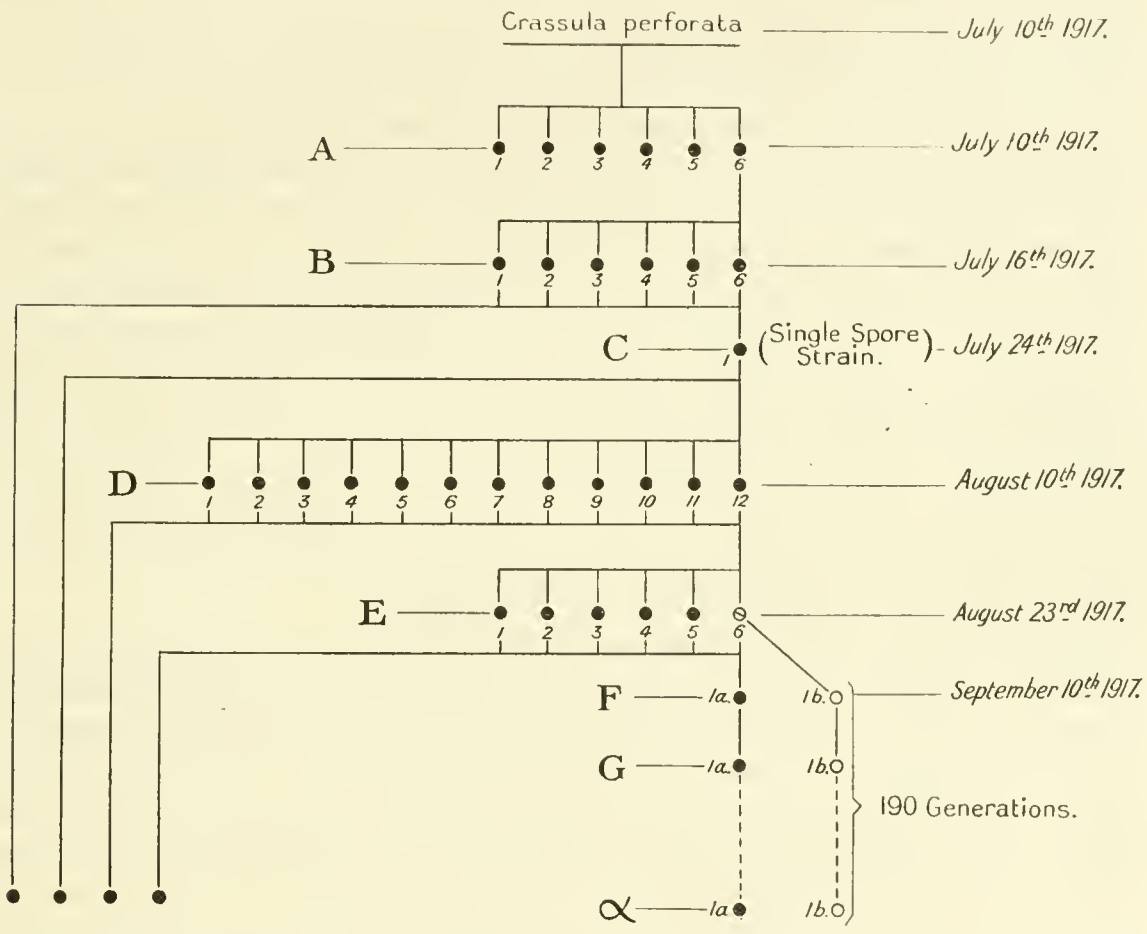
### 3. Origin of Colourless Form. General.

On July 10, 1917, a diseased specimen of *Crassula perforata* from the succulent house of the Royal Botanic Gardens, Kew, was examined, and the causal pathogen identified as *Botrytis cinerea*. The growth was normal in every respect, and from it six cultures, (A. 1–6), were prepared. These appeared to be free from contamination by other organisms, and on July 16, six further cultures, (B. 1–6), were taken from (A. 6). On July 22 all twelve were carefully examined and found to be free from contamination. A series of dilution drops were then prepared from (B. 6), many of these drops containing but a single spore. The latter were transferred to a thin plate of potato-agar and their development scrutinised carefully during two days. A colony whose origin from a single spore had been ascertained with certainty was then picked up on a platinum wire and planted in a tube slant. The spores are comparatively large, the most common size on the host plant being $7{\cdot}5\ \mu \times 9{\cdot}5\ \mu$; they are thus easily seen and isolated. On germination they become still more obvious, and the obtaining of a single-spore strain is thus a simple procedure. On August 10 twelve cultures (D. 1–12), were prepared from the pedigree stock (C. 1), and on August 23 six further cultures, (E. 1–6) were made from (D. 12). So far the work had merely been a part of the ordinary laboratory routine, the cultures which were all on tube slants of potato-agar being used for various experimental purposes.

About a week later cultures (E. 1–6) were examined and some three or four days later re-examined. The six cultures had all been placed together in a wire basket and left on the laboratory bench. It was noted that in tube (E. 6) one of the sclerotia had not turned black like the others but had remained colourless. In certain strains of *Botrytis cinerea* these bodies do not form the black pigment characteristic of the genus until they are mature, and although this particular sclerotium differed from its fellows in apparently following this mode of behaviour, it was merely regarded as a somewhat youthful individual, perhaps a little unusual, but not particularly noteworthy. On September 10, however, when tubes (E. 1–6) were again examined and the aberrant sclerotium in (E. 6), now quite mature, was found to be still lacking any sign of pigmentation, it was decided to isolate this sclerotium and breed from it. In tube (E. 6) there were in addition to the colourless sclerotium 24 normal black sclerotia and a black sclerotial crust around the bottom of the tube, where the medium had contracted away from the glass.

The colourless sclerotium was removed and thoroughly washed in sterile distilled water to remove adherent spores. It was then broken open and a fragment from the centre taken out, placed on a tube-slant of potato-agar (F.I. *b*) and incubated at 22° C. A vigorous growth ensued and a week later sclerotia began to be formed. In the tube there were ultimately formed 17 sclerotia and all were colourless.

TABLE I.




In a normal sclerotium the pigmentation is confined to the outer one to three layers of cells (text-fig. 1) and as it is from these that the conidiophores arise on germination, it was considered not impossible that these cells and their derivatives only might carry the power of colour formation. If, therefore, a fragment of the inner hyaline tissues of such a normal sclerotium were planted on a nutrient medium it seemed possible that colourless sclerotia similar to those obtained in tube (F.I. *b*) might develop. Accordingly, a black sclerotium from (E. 6) was treated thus—exactly as the aberrant sclerotium had been treated. The subsequent culture (F.I. *a*) contained uniformly black sclerotia.

There were now in existence two distinct strains (F.I. *a*) and (F.I. *b*) of the fungus

both derived from a single-spore stock (C. 1). Of these the parental form (F.I. *a*) possessed the highly characteristic black sclerotia; whilst the new strain (F.I. *b*) which had arisen from this apparently by a single direct saltation, possessed colourless sclerotia. The history of the two forms is shown in a diagrammatic manner in Table I. The distinguishing lettering (A. 1–6) and so forth, and the arrangement of descent in this Table have been adopted for the sake of clearness in the account. In practice (D. 1–12) were merely 12 similar cultures taken from (C. 1), and (E. 6) was taken from one of these cultures chosen at random. In the same way six cultures were made on July 16 and from one of these chosen at random (C. 1) was prepared.

It seemed possible that an albinistic tendency might be a "weakness" inherent in this particular race of *Botrytis cinerea*, and that if the preceding generations were further subcultured colourless sclerotia might again appear. Unfortunately, the original diseased plant and the first six cultures made from it had been destroyed, but all subsequent cultures had been retained. As no particular care had been taken of these some had become contaminated. In all cases, however, it was possible to subculture purely from them and this was done extensively for many generations under the most diverse conditions of light, temperature, food supply and so forth, all the resulting cultures being most carefully examined for any indication of a lack of sclerotial pigmentation. All were perfectly normal.

The parental strain and its colourless derivative have each, up to the time of writing, passed through over two hundred direct tube generations with very many lateral subcultures, and although these have been placed under diverse environmental conditions each strain has remained absolutely constant.

Since this colourless strain arose, many thousands of cultures and fresh specimens of *Botrytis cinerea* have passed under my observation, but in no single instance have I found any similar absence of sclerotial pigmentation. Prof. H. H. Whetzel, of Cornell University, who for some years has been engaged upon a monograph of the genus Botrytis and has received and compared material from many parts of the world, informs me, after having examined a culture of the colourless strain, that he has "never seen anything quite like it." In the voluminous literature devoted to the fungus there is no record of a form with colourless sclerotia.

Having regard, therefore, to our knowledge of the fungus *Botrytis cinerea* and to the particular conditions under which the aberrant strain originated, it would appear difficult on present criteria to interpret it as other than a true mutation.

### 4. Comparison of Parental Strain and its Colourless Derivative.

The obvious difference between the original and the derived forms lies in the colour of the sclerotia. Mutations in the fungi *Aspergillus niger* and *Penicillium glaucum* have been described at length by Arcichovskij (2), Waterman (83) and Schiemann (69). In these cases, however, although the obvious change was in spore

colouration, there were other less visible changes in several independent respects, certain morphological characters and physiological activities of the derived fungi being different from those of the original forms. In addition to these facts, the results of recent researches on the linkage of characters and the multiple allelomorph interpretation of quantitative inheritance made it appear probable that in the present case the colour change might be only one of many related changes less obviously visible.

Moreover, whether one regards the black sclerotial pigment as a mere excreted product or as an integral and actively functioning substance in the metabolic activities of the organism, the pigment itself is the result of a long and elaborate series of causally dependent processes, and a sudden and permanent loss of the power to excrete or form such matter can only be the visible expression of deep-seated physiological changes in the developing organism. But a living organism is an extremely complex and delicate equilibration of rhythmic metabolic processes in a colloid substratum, and it is difficult to conceive of any sudden derangement of these activities which has not reverberations in many directions.

It seemed, therefore, very desirable to make some comparison of the parental strain and its colourless derivative to ascertain the nature of any changes other than pigmentation which might be present. This comparison proceeded along several lines, the main directions of which are indicated briefly below.

### A. *Physiological Activities.*

If, as evidence would seem to show, the physiological reaction of a particular organism is constant only in an unvarying environment, it follows that identity of response to like stimuli implies identity of constitution, and, conversely, that a change in physiological constitution will be reflected in a changed reaction to unaltered conditions. Only if this is true may a comparison of the physiological activities of two organisms be instituted; and being true, it will afford the most delicate test of individual or genotypic identity, for the technique is quantitative and the nicest differences may be measured.

(a) *Germination of Spores.*—The effect of the character of the nutrient medium upon the germination of fungus spores is often very marked, and not infrequently this relation may be used to differentiate two or more genotypes otherwise difficult to distinguish. The percentage germination within certain periods of time when the spores were immersed in distilled water, 15 per cent. gelatine in water, and Coons' solution (21) were compared in the two experimental forms and in two strains used as a control. The spores were taken from three-day cultures on potato-agar incubated at 22° C., and the hanging drops in which the germinations were tested were maintained in the cool incubator at 18° C. The results are shown below, the numbers representing percentage germinations.

| Medium. | Parental Strain. | | | Colourless Strain. | | | Control Strain A. | | | Control Strain B. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 6 hours. | 12 hours. | 24 hours. | 6 hours. | 12 hours. | 24 hours. | 6 hours. | 12 hours. | 24 hours. | 6 hours. | 12 hours. | 24 hours. |
| Distilled water | 10 | 27 | 55 | 12 | 25 | 53 | 33 | 51 | 68 | 25 | 55 | 93 |
| 15 per cent. gelatine | 43 | 67 | 93 | 41 | 66 | 90 | 47 | 78 | 90 | 47 | 72 | 90 |
| Coons' solution | 63 | 83 | 97 | 61 | 80 | 96 | 63 | 90 | 98 | 61 | 72 | 98 |

The lethal temperature of the spores was tested by making spore suspensions in 15 per cent. gelatine in distilled water and heating corresponding tubes of parental, derived, and control strains together in a water bath for ten minutes, the temperature rising by increments of five degrees. The experimental strains both gave growth in tubes heated to 45° C., but not in those heated to 50° C., whilst in the control tubes growth was inhibited by temperatures of 45° C. and 55° C. respectively.

(b) *Growth on Nutrient Media.*—One of the most critical and easily determined measures of the identity of two fungi is their growth and behaviour upon various standardised nutrient media under controlled conditions. The general characters of the colony, the colour reactions of hyphæ and medium, and the many other detailed phenomena which may be observed are valuable diagnostic characters, and sharply reflect the physiological differences separating two genotypes. At the end of his paper on "Cultural Studies of Species of *Penicillium*," Thom (77) has drawn up a scheme of comparative cultural data, which, in the present absence of any more detailed formulation, could with valuable results be adopted by mycologists as a basis of systematic diagnosis. In the comparative examination of the experimental strains this technique was followed, and the results compared with those of two control strains. No essential difference could be detected between the former, but these stood in marked contrast with the latter. For laboratory purposes other than this particular investigation, the parental strain and its colourless derivative were largely used in experiment, so that the cultural comparison of the two forms was amplified in very many ways. In all cases where they were grown under parallel conditions no essential difference could be detected. The comparative development of the two strains and control cultures was tested further at various temperatures, and with different intensities of light and conditions of aeration, but the results obtained only confirmed those referred to above.

A few qualitative experiments were carried out to test the comparative enzyme production of the fungi, the methods used being adapted from those described by Crabill and Reed (22) in 1915. On starch-agar the experimental strain and one control showed approximately equal amylolytic action under the colonies, while the second control showed a distinct halo indicating the presence of a diffusible extra-

cellular amylase. These differences were thrown up more clearly when the plates were flooded with iodine solution. On litmus-cream-agar growth of the experimental strains was profuse, and there was strong acid production, the red colour diffusing through the medium. In both control strains the lipolytic action was equally marked but the redness much more sharply defined. All the four strains liquefy gelatine and slowly dissolve fibrin in fibrin-agar. In both experimental strains the lipase production is feeble and only slight browning of the fibrin occurred, this being markedly different from the results in the controls. On casein-agar moderate growth occurred, and ereptic action was prominent, extending in a distinct band around the colonies. The production of erepsin was distinctly greater in both controls than in the experimental strains. Sparse development occurred on asparagin-rosolic-acid-agar with slight amidase production. The red colour in the experimental strains was less diffuse than in the controls.

Cytase formation was tested by growing the organisms in Dox's solution (23) the carbon source being filter paper. Moderate growth occurred, but after a fortnight there was only a barely perceptible reduction with Fehling's solution in the experimental strains and one control strain, whilst the second control showed more distinct cytolytic action. Calcium-carbonate-agar prepared with Dox's solution (23) plus one per cent. of lactose showed a feeble though distinct production of lactic acid, this being more marked in the experimental than in the control strains. In all the experiments no essential differences in the enzymic activities of the normal strain and its aberrant form could be detected.

(c) *Pathogenicity.*—Marked and constant differences in virulence for different hosts are shown by the various strains of *Botrytis cinerea*. In testing the comparative pathogenicity of the experimental strains, the spores, together with a fragment of potato-agar, were planted on corresponding surfaces such as the opposite sides of a fruit, or opposite halves of a leaf. On banana, apple, and tomato fruits, lily and potato leaves, tulip bulbs, and twigs of horse-chestnut infection readily occurred, and vigorous development ensued. On onion-bulb scales, potato tubers, crab-apple, cucumber, and fruits of *Pyrus japonica*, leaves of *Primula sinensis* and wall-flower, infection occurred with very considerable difficulty and only when the tissues were badly bruised, and the growth tended to die out rapidly. Bulbs of snowdrop and leaves of Portuguese laurel and rhododendron could not be infected. In no case was any essential difference in pathogenicity noted between the parental and colourless strains. Markedly different results were obtained with the two control strains, of which one was originally derived from onion bulbs and the other from fruits of *Pyrus japonica*.

The cultural examination of the parental strain and its colourless derivative, although put separately above, was only part of a much larger series of experiments which were being carried out in the laboratory to ascertain the physiological differences of closely similar morphological strains, and the identity of result obtained

in this comparison very largely takes its critical value from its relation to this wider issue. Briefly, it may be stated that so far as the physiological constitutions of the parental strain and the colourless strain may be judged from their behaviour, the two strains are identical.

B. *Morphological Characters.*

Phenotypic plasticity is shown by most fungi, and is not only evident in those general characters which together constitute the morphological facies of the organism, but in those critical structures, the reproductive bodies, whose constancy has so often been assumed.

(d) *Dimensions of Spores.*—During recent years it has been shown in a general way by many students of the fungi, notably perhaps by Stevens and Hall (76), and following their investigations by Elliott (26), Moreau (53), Mutto and Pollacci (57), Gäumann (30, 31), and others, that the size of fungus spores is a function of the particular organism and the environmental conditions. With regard to the present fungus, *Botrytis cinerea*, under certain standardised conditions, the spore dimensions are a function of the particular pure line and the quality of the nutrient substratum ; and for any particular pure line, different standardised food-media give different but mathematically constant quantities for the modal values of the variation curves of the spores. On a series of standardised media under constant conditions, therefore, the modal values of the spores of any pure line of the fungus may be represented by a curve which is constant and characteristic for that pure line. The critical quantity is not the extremes of variation, but the curve of modal values ; and, other factors not varying, a difference of genotype is immediately reflected in this curve of modal values.

The modal values of the spores of the parental strain and its colourless derivative were compared in this way, the number of measurements for each particular test being 500. The results of three such comparisons are shown on p. 93.

The conidiophores of the fungus are comparatively large and complexly branched structures, and the mechanical difficulties in the way of their minute comparison are such as to render it difficult to eliminate errors of selection. In so far, however, as these structures could be compared with regard to the length, diameter, manner of branching and abstriction of spores no differences could be distinguished between the normal strain and its aberrant form ; and this holds true also for the colour of the spore mass.

(e) *Haptera.*—One of the most interesting structural features of *Botrytis cinerea* are the organs of attachment, which were described by de Bary (5, 6) under the name "Haft-organen," and are variously termed "haptera" or "appressoria."

Their presence or absence in a culture is determined by environmental conditions, and the controlling factors are different for the several strains of the fungus. The experimental strains behaved similarly in this respect, forming few haptera when

| Medium. | Spore. | — | Parent. | Derivative. |
|---|---|---|---|---|
| Potato-agar | Length | Minimum | 6·0 u | 6·0 u |
| | | Mode | 9·5 u | 9·5 u |
| | | Maximum | 12·0 u | 12·0 u |
| | Breadth | Minimum | 4·5 u | 5·0 u |
| | | Mode | 7·5 u | 7·5 u |
| | | Maximum | 10·0 u | 10·0 u |
| Steamed potato | Length | Minimum | 6·5 u | 6·5 u |
| | | Mode | 10·5 u | 10·5 u |
| | | Maximum | 16·0 u | 16·5 u |
| | Breadth | Minimum | 5·5 u | 5·5 u |
| | | Mode | 9·0 u | 9·0 u |
| | | Maximum | 13·5 u | 14·5 u |
| Czapek's medium | Length | Minimum | 6·0 u | 6·0 u |
| | | Mode | 11·5 u | 11·5 u |
| | | Maximum | 17·0 u | 17·0 u |
| | Breadth | Minimum | 5·5 u | 5·0 u |
| | | Mode | 9·5 u | 9·5 u |
| | | Maximum | 14·0 u | 14·5 u |

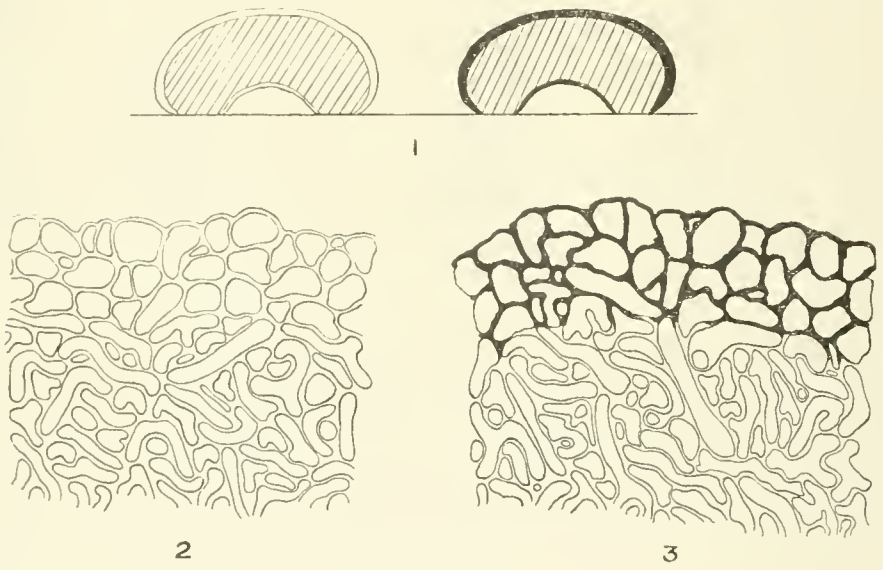
incubated at 22° C. on CZAPEK'S medium lacking carbon, producing them in abundance when M/10 glucose is added to this medium, and suppressing their formation when M/1 glucose is added. A similar course of behaviour is evident when the substratum remains constant and the temperature varies. The particular relations between hapteral formation and controllable external factors are very individual for the several strains of *Botrytis cinerea*, and constitute a diagnostic feature. The identity of behaviour between the normal strain and the colourless strain is therefore noteworthy.

(f) *Sclerotia.*—The type of correlation between hapteral formation and environment which has been indicated above holds true in the formation of sclerotia, and the presence or absence of these in any particular strain may be determined at will. It may briefly be stated that under all circumstances the modes of behaviour of the two strains with respect to this structure were alike.

Owing to the widely held belief regarding the constitutional weakness of albinos, as compared with the normal type, it was considered not improbable that the colourless sclerotia might show a lack of vitality and power to withstand adverse conditions. No indication of such weakness had appeared in the general cultural comparison of the two strains, but it seemed desirable to test the sclerotia themselves, for it is a common assumption that dark coloured or "carbonised" hyphæ are more resistant than hyaline hyphæ. The comparison was made by testing the relative germinative capacity of selected black and colourless sclerotia after subjection to various treatments. Parallel series exposed to extremes of temperature and desiccation for various periods exhibited no differential resistance, nor was this found on subjecting

the sclerotia to saturated atmospheres of chloroform and ether, or various dilutions of toxic substances such as mercury bichloride, copper sulphate, carbolic acid, and picric acid. Other strains used as controls gave markedly different results.

In structure the sclerotia vary according to the conditions under which they are formed. In contact with a hard surface they show an outer compact zone two or three cells in thickness which merges into an inner zone consisting of a firm colourless tissue-mass formed by closely interweaving hyphæ (text-fig. 1). On a soft matrix a third central zone of loose spongy hyphæ is usually present. The latter type of sclerotia are more or less spherical in shape whilst the former often fuse




TEXT-FIG. 1.

1. Vertical sections of colourless and normal sclerotia. Diagrammatic.
2. Vertical section of colourless sclerotium.
3. Vertical section of normal sclerotium.

} Swift ⅓ obj. No. 4 eyepiece.

together into a thin concave crust. The contact surfaces of both forms are colourless, but this is merely due to lack of aeration, for if the sclerotia be removed so that all surfaces have access to air, the previously colourless contact surfaces develop the tissue containing the characteristic pigment. If, however, the sclerotia be cut across so that the inner tissue zones are exposed, these do not develop pigment. The power of colour formation is therefore confined to the peripheral cells. Save for the absence of pigment in these cells, the most minute examination of the structure and form of the sclerotia of the parental strain and its aberrant form failed to reveal any difference.

Like the physiological examination the comparison of the morphological characters of the two forms under discussion has greatly increased value when it is realised that this work was only part of a wider series of experimental researches, which functioned as controls, and the results of which were invaluable in aiding in the

interpretation of doubtful phenomena. The conclusion drawn from the entire study was that the parental strain and the aberrant derivative differed in respect only to the one character of sclerotial pigmentation. The phenomenon would thus appear to form an exact parallel with mutations by loss of single genes in the higher organisms, and it would stand in contrast with the fungal mutations described by Arcichovskij (2), Schiemann (69), and Waterman (83), in which there was simultaneous variation in several independent respects.

The explanation of an apparent genotypic alteration on the basis of such a profound change as the sheer "dropping out" of an element from a most delicately balanced reaction-system would appear out of the question. I can only visualise the organism as a unity in itself, in the sense that it consists of a very great number of elements, each an elaborate reaction-system, which bear a specific and causal relationship to one another. The continued harmonious functioning of the total system minus one of its parts is to me unthinkable.

Pure lines of organisms are often separated by characters which are extremely minute and difficult to detect, as for example the obscure serological reactions which differentiate certain races of bacteria, and it would appear probable that were the analysis in the present case carried to a further degree of refinement, many deeply underlying and elusive physiological differences between the two strains would be detected. Such work, however, would constitute a most laborious and intensive study quite beyond the scope of the present paper.

### 5. Origin of Colourless Form. Special.

A question of some considerable importance is the exact point of origin of the colourless strain. As described it apparently arose in the single colourless sclerotium found in tube (E. 6), and for long no doubt was entertained that this was its actual point of inception. In the absence of further evidence, however, there was no certainty of this, for the colourless strain would only be observed on the production by it of sclerotia, and if sclerotia were not formed the presence of this strain could not even be suspected.

Cultures (D. 1–12) had been made from the pedigree stock (C. 1) in the usual way, by transference of spores on a platinum wire, and a large and indefinite number of these reproductive bodies would thus be implanted in each tube.

Approximately 100 per cent. of these spores would germinate, and the resulting growth would therefore comprise very many intermingled mycelia, some or all of which would produce conidiophores and spores. Similarly, the six tubes (E. 1–6) of August 13, prepared in the same way from one (D. 6) of the previous twelve, would each contain an indefinite number of mycelia, each mycelium corresponding to a single individual. Each tube might, therefore, contain one hundred or more individuals of identical genotypic constitution, growing inextricably mixed together and all or few producing spores and later sclerotia.

Furthermore, the first crop of spores produced from the mycelia in the tube may fall to the surface of the medium, immediately germinate, and two or three days later again reproduce, when the same process may be repeated. Thus in each *test-tube* generation there may be two or more *lineal* generations, depending partly upon the conditions to which the growth is subjected, and partly upon the available germinating surface of the medium.

From the pedigree stock (C. 1) to the culture of August 13 (E. 6) containing the first colourless sclerotium there were present three direct test-tube generations, which may actually consist of from about six to twenty lineal generations, and at some one point in this series of life cycles the aberrant strain may have arisen. In the series of cultures under consideration there was no evidence that any accessory spore form was produced, nor were the conditions such as would be conducive to their formation. A brief discussion of the two possible points of origin, the spore and the sclerotium, is perhaps desirable.

### C. *Spore Formation.*

It will be clear that the colourless strain may have arisen as a variation in a single spore at any point between the second lineal generation in tube (C. 1) and the last lineal generation in tube (E. 6). If, for example, in the former, none of the offspring of this spore, although present in the later cultures, produced sclerotia until a favourable opportunity occurred in tube (E. 6). As, for essential purposes, the colourless form differs from the parental form only in the absence of sclerotial pigmentation, the presence of the former could not possibly be ascertained in a culture of the latter in the absence of the formation of its sclerotia. If, therefore, the aberrant form arose as a single spore variation, it is impossible to locate exactly its point of origin.

Certain facts relevant to the foregoing must, however, be noted. It has been an invariable technique throughout the whole of the research, when transferring inoculum from one tube to another, to take this either from a single conidiophore or, preferably, from a single germinated sclerotium. Now, every conidiophore and each sclerotium arises not only from a single original spore, but actually takes origin in a single mycelial cell. It will be obvious that this very considerably narrows the range wherein the aberrant form may have arisen as a spore saltation.

Furthermore, from its inception the colourless strain has been constantly reproduced by means of spores, and if it arose as a spore variation, the very existence of the new form implies that this particular spore germinated and produced a mycelium giving rise to conidiophores and spores, the latter possessing the new potentiality. But a single mycelium gives rise to an immense number of conidiophores, and the number of spores produced on even a single conidiophore is very great. It is, therefore, almost unthinkable that the new character should not again have appeared in any of the direct and lateral cultures made from tube (E. 6) and from the previous generations.

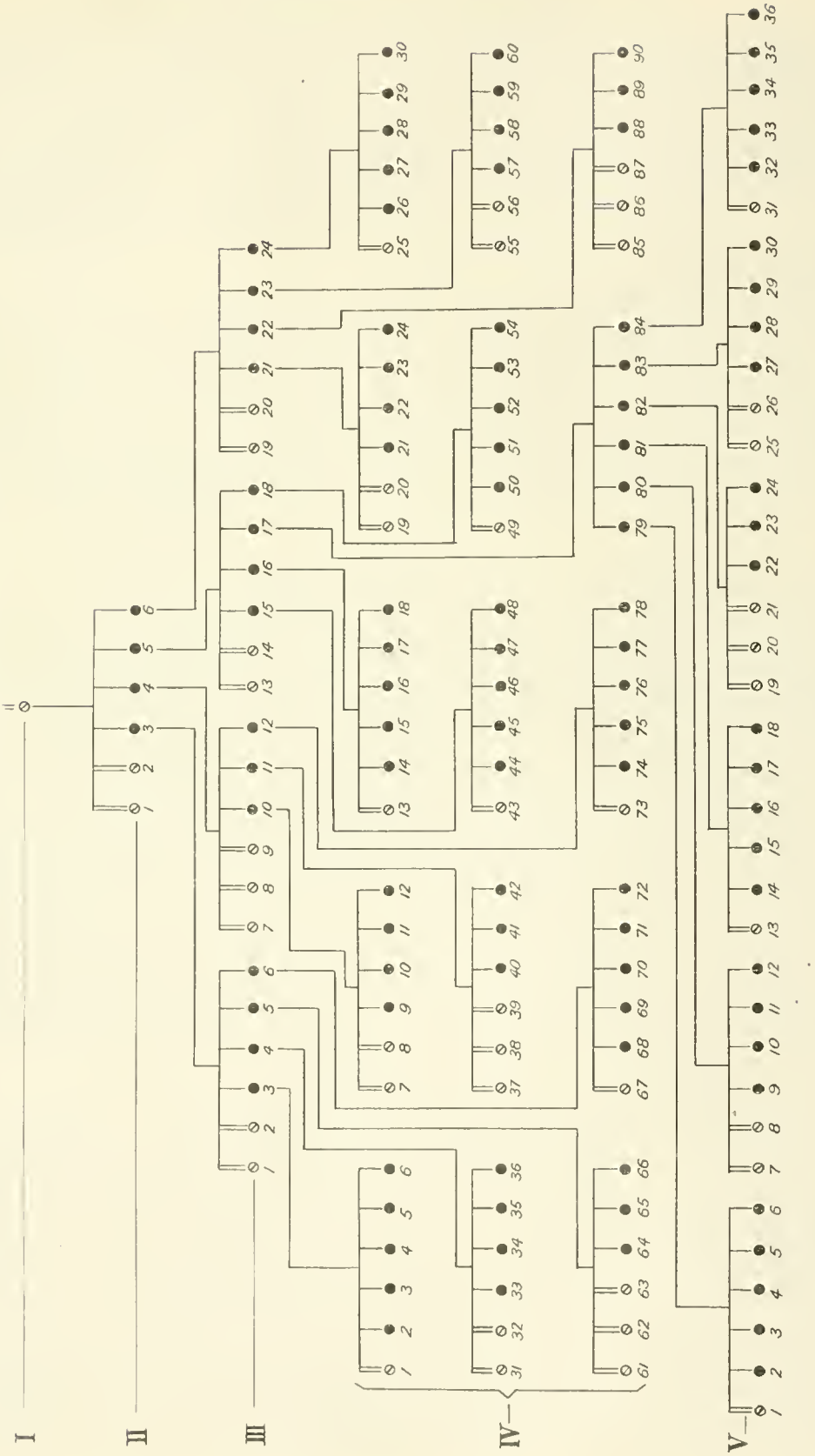
For purposes other than this particular piece of investigation it was decided to test the probability of the "throwing up" of the colourless strain in a parental culture, when only a single spore of the former is present. Accordingly, inoculum from (E. 4) was transferred to a clean tube-slant and streaked over the surface, after which there was inserted a single spore of the colourless strain. The circumstances would thus correspond in all essentials to those in tube (E. 6) if the aberrant form had arisen as a variation in a single spore. The results arranged diagrammatically are shown in Table II.

Tube (I) was prepared on March 2 and ten days later, when the sclerotia were mature, it was found that of the 23 sclerotia in the culture 2 were colourless, the aberrant strain thus "throwing up" in the first tube generation. On March 9 cultures (II. 1–6) were taken from (I) and ten days later it was found that 2 of these contained sclerotia of both strains, the remaining 4 containing only black sclerotia. In (II. 1) there were 15 black and 1 colourless sclerotium, and in (II. 2) 27 black and 1 colourless sclerotium. From each of the tubes containing only black sclerotia (II. 3–6) six cultures were made (III. 1–24). Of these 24 cultures 9 contained both parent and colourless form and the remainder black sclerotia only. In the former there were 162 black and 17 colourless sclerotia. From each of the 15 tubes containing only black sclerotia 6 cultures (IV. 1–90) were prepared. Of these, 24 tubes contained both black and colourless sclerotia in the proportion of 528 of the former to 39 of the latter, and the remaining 66 tubes contained only black sclerotia. In the 6 tubes (IV. 79–84), derived from (III. 17), only black sclerotia were present, and it appeared that in this line the colourless strain had "run out." To test if this were so, 6 cultures (V. 1–36) were made from each of these tubes. In at least 1 tube in each set of 6 the colourless strain was again thrown up, there being in all 10 tubes containing both kinds of sclerotia and 26 tubes containing only the black form. The work was not carried further than this; but in the total experiment, which represented 5 tube generations and 157 cultures, the aberrant form appeared in 46 tubes. A parallel control series prepared with similar inoculum from (E. 4) was synchronised with the foregoing, similar in all respects save that the single spore of the colourless strain was omitted. All the sclerotia were black.

It will be clear from this experiment that the presence of a single spore of the aberrant strain in a parental tube may cause the new character to appear in the first generation, and in a large proportion of cultures of succeeding generations. As already stated, the generations (C. 1), (D. 1–12), and (E. 1–6) were all subcultured freely, but the aberrant strain was absent from all descendants. In nearly 200 direct tube generations of the parent strain (F. 1*a*) the aberrant form has not appeared. This can only be explained by the hypothesis that the colourless strain did not arise as a spore variation, but arose in and was confined to the initial colourless sclerotium, and that the removal of this eliminated the aberrant strain from the parental culture (E. 6). The new strain was thus seen immediately on its inception and isolated.

TABLE II.

Inoculum from (E. 4) + one Spore from Colourless Strain.



### D. *Sclerotial Formation.*

It will be evident from text-fig. 1 that the sclerotium consists of a mass of more or less firm tissue, enclosed in a thin layer of cells, in the walls of which the black pigment is located. Now, unless the sclerotium arises in a single hyphal cell, in which the variation may have occurred, the physiological change resulting in loss of colour must have taken place simultaneously in all those peripheral cells which normally would form pigment. In view of the fact that the new character is strictly hereditary, and obtains by the growth of any portion of a single individual, and that the initial colourless sclerotium arose in a tube culture containing a perfectly normal growth, such a deep-seated physiological change occurring simultaneously in a very great number of cells, and to an identical degree in all, is a process extremely difficult to conceive.

Very little is known of the initial developmental stages in the formation of sclerotia, but the latter are usually considered to be a further growth or modification of haptera, structures which are very accurately delineated in the works of Marshall Ward (80) and Istvanffi (33). In their formation one, or more usually several related hyphæ form short branches on which arise tufts of secondary branches. These are directed vertically to the contact surface, and the clustered hyphæ mat together laterally forming a solid body of a tassel-like shape and structure.

It is of interest to note that de Bary (5, 6) regarded sclerotia and haptera as discrete bodies, stating that although clusters of haptera "have been mistaken for sclerotia . . . they have no connection." Munn (56), who has most recently investigated this fungus, writes: "It was observed that in many cases the appressoria are incipient stages of the sclerotia (Plate 3, figs. 23, 24)." Now if Munn's figs. 23, 24, and 13, which represent stages in the development of sclerotia, be compared with his figs. 25 and 27, which show the development of haptera, the essential difference is plain; and this distinction is obvious in the many delineations of these structures by previous investigators. There are no transition stages. The general laboratory study of the fungus and the relation of hapteral and sclerotial formation to various media and known conditions show that the structures are organs *sui generis*. This was confirmed in a very striking way by the development of these bodies in the aberrant strain. The haptera of the normal strain very rapidly become dark coloured and finally almost black, and this similarity of colouring with the sclerotia is in large part responsible for the confusion. In the case of the colourless strain, however, it is most helpful, for whilst the sclerotia have completely lost their pigment, the haptera have retained it, and thus in the one culture there co-exist the albinistic sclerotia and brown or olive-black haptera. This colour difference is distinctive from the inception of the two structures.

If therefore the sclerotia and haptera are discrete entities, the earliest stages that we know in the development of the former are the clusters of interweaving hyphæ

depicted by Istvanffi, Munn, and others, and there is no reason why these should not have arisen in the branching of a single mother-cell. On searching selected cultures both by sectioning, and more successfully by mechanical teasing and *intra vitam* staining with neutral-red and methylene-blue, the initial stages were soon found. For the present issue, the main point of interest is that each sclerotium arises symphogenetically from an individual cell of the mycelium.*

On the open surface of a soft matrix the sclerotia usually remain separate, but if developed in a fissure or along a line of contact they often fuse to form a continuous crust. Figs. 50 and 51 on Plate 15 of Istvanffi's memoir (33) are to be interpreted in this latter way. If mixed inocula of parental and colourless strains be grown, interesting composite sclerotial crusts may very rarely be obtained. In these there is no intimate mingling of hyphæ, but merely a lateral "grafting," and if such a composite sclerotium be germinated, the conidia and conidiophores, although indistinguishable, each reproduce purely their original parental type.

From the foregoing discussion it will be clear that the physiological change which resulted in the production of the aberrant form must have occurred in the mother-cell which gave rise to the initial colourless sclerotium in tube (E. 6).

### 6. Nature of Loss of Colour.

A detailed biochemical study of sclerotial pigmentation is beyond the scope of the present paper, but one may perhaps refer to certain simple experiments which throw some light upon the nature of the physiological change which has occasioned the production of the colourless form.

During the last two decades the pigmentation processes in animals and plants have been the subject of much investigation, and although this has followed the two distinct lines of biochemistry and genetics, the results have complemented each other in a quite remarkable way. The general trend of evidence would appear to lead to the view that the pigmentation processes throughout are essentially of an oxidatory nature with associated subsidiary or perhaps preliminary reduction processes.

Albinos have been shown to be divisible into two categories termed "dominant whites" and "recessive whites" respectively. In the latter, the absence of pigment is due to the lack of either the oxidase factor, or the chromogen factor, or of both; whilst in the former, both oxidase and chromogen are present, but their interaction is prevented by a third inhibiting factor. The oxidatory nature of the pigmentation processes is common to so many widely diverse organisms, that one might with some justification expect to find it even in the fungi, and consequently to find also that albinism in such an organism as *Botrytis cinerea* showed common features with the same phenomenon in rabbits or primulas. This fungus, however, is asexual, and the terms "dominant" and "recessive" are inapplicable; but if the pigmentation

* In certain strains of *Botrytis cinerea* the sclerotia originate meristogenetically.

processes are parallel to those of other organisms, the physiological changes which have resulted in the production of the colourless form may involve either the presence of a new inhibiting factor or the loss of either oxidase or chromogen or both.

The simplest and most direct method of examining the condition of the aberrant form is to test the colourless sclerotia for the presence of oxidase. If this be present and active it follows that the chromogen factor must be absent for otherwise pigmentation would have ensued. If there is no oxidase reaction, then either this factor is lacking or its activity inhibited by a third factor. The technique was based on that devised by KEEBLE and ARMSTRONG (41, 42, 43) in their investigations of flower pigments. Solutions of benzidine, $\alpha$-naphthol, and $p$-phenylenediamine gave a marked reaction, whilst with a solution of guaiacum-resin a barely perceptible lavender-blue colour was obtained.

These reactions were unaltered by preliminary treatment of the sclerotia with absolute alcohol or hydrocyanic acid, or by the subsequent addition of hydrogen-peroxide, and in all cases they were produced in a few minutes irrespective of temperature differences ranging from that of the laboratory to incubation at 37° C.

The distribution of the oxidase in the colourless sclerotia was identical in every case, and was noteworthy, being confined to the walls of the peripheral cells which in the parental form contain the normal black pigment. So exact was this agreement that some of the sections of colourless sclerotia treated with benzidine were barely distinguishable from sections of normal sclerotia.

Whether the colourless sclerotia have been formed in light or darkness makes no difference to the intensity of the oxidase reaction, nor is this changed by growth of the fungus at various temperatures or upon nutrient substrata presenting very different physiological conditions.

On WHELDALE'S view (85) the chromogens are flavones or flavonal glucosides. The presence of these bodies in tissues may readily be detected by the bright colour given on exposure to ammonia vapour, and the dull-green (or sometimes red-brown) colour with ferric-chloride. When these tests were applied to sections and pulps of the colourless sclerotia a totally negative result was obtained.

NEILSON JONES (39) has shown that even in certain recessive whites both chromogen and oxidase appear to be present, but that owing perhaps to their discrete distribution interaction is prevented. To test the possibility of this explanation of the albinism of the aberrant strain a number of sclerotia were boiled in 5 c.c. of 50 per cent. alcohol to destroy the oxidase, pulped, filtered and the filtrate evaporated to a few drops. Sections of the colourless sclerotia were then immersed in this and incubated at 27° C. No trace of oxidase reaction was observed, confirming the absence of chromogen in the sclerotia.

True brown and black pigments are of the very greatest rarity in the higher plants. MÖBIUS (52) however has shown that they do exist as in the spots on the alæ of Vicia Faba, but even here the colouring matter is dissolved in the cell-sap.

In thin sections of the normal sclerotia of *Botrytis cinerea* the pigment appears as a true brown-black. Furthermore, as shown in text-fig. 1, it is contained in the substance of the cell-walls and not in the cell-content which is perfectly hyaline. It thus differs very markedly from the usual plant pigments, and would appear to be more closely related to the melanotic pigments, certain of the melano-proteins existing dissolved in the keratin structures.

The mother substance of melanin is generally supposed to be tyrosin, this colourless chromogen being converted into a black pigment by the oxidising enzyme tyrosinase. If, therefore, the black pigment in the normal sclerotia be melanin, the aberrant sclerotia which presumably lack the factor for tyrosin should exhibit a tyrosinase reaction. The tests for oxidase already referred to are only for the detection of phenolase or laccase which Bourquelot and Bertrand (14), Zellner (88), Pringsheim (63), Kastle (40) and others have shown to be almost universally present in fungi.

Tyrosinase is comparatively rare in its distribution in fungi. Bertrand (11), however, has shown its presence in species of *Russula*, especially *Russula nigricans*, and Lutz (49) in the stipes and pilei of *Gyromitra gigas* and *Disciotis parlata*; while more recently it has been demonstrated in *Lenzites sæpiaria* by Zeller (87) and in various species of *Actinomyces* by Drechsler (24).

According to Bourquelot and Bertrand (14) the best substrate for tyrosinase is tyrosin. Colourless sclerotia incubated at 18° C., 25° C., and 37° C. for 14 days in a saturated aqueous solution of silk-tyrosin gave a totally negative reaction, nor could this be changed by preliminary treatment with alcohol, ferrous sulphate or hydrocyanic acid or by subsequent treatment with hydrogen peroxide. A pulp of sclerotia, or sections, gave a similar result with this treatment.

Chodat and Staub (20) have advocated the use of *p*-cresol with the addition of glycin for the detection of the enzyme tyrosinase. Sclerotia immersed in this and incubated at 37° C. showed after many days a diffuse pale-yellow tint instead of the rich violet-blue reaction characteristic of tyrosinase.

In 1907 Abderhalden and Guggenheim (1) showed that N/100 hydrochloric acid permanently destroyed the activity of tyrosinase, but in the present case an immediate and direct oxidase reaction with benzidine was obtained after preliminary treatment of the colourless sclerotium with N/40 hydrochloric acid, although a strength of N/30 destroyed its activity.

It was found by Bertrand (11) that tyrosinase could be differentiated from other oxidising enzymes by its lower lethal temperature, its activity being inhibited at 50° C. The oxidase in the colourless sclerotia is inhibited only by a temperature of from 80° C. to 85° C.

Furthermore Bach (3, 4) has shown that the pigment formed by the action of tyrosinase upon tyrosin may be oxidised further to a colourless compound by a dilute acid solution of potassium permanganate. The normal black sclerotia of *Botrytis*

*cinerea* remained unaltered when treated in this way. The black pigment is indeed extraordinarily stable, remaining apparently unchanged after boiling in concentrated acids and alkalies. When treated thus the stuff of the tissues disappears whilst the pigment remains as a most delicate gossamer-like cellular film. This stability immediately distinguishes the sclerotial pigment of *Botrytis* from other black fungal pigments such as the Aspergillin of *Aspergillus niger* which is soluble in hot water, or the colouring matter of *Daldinia*, *Hypoxylon* and various other carbonaceous Ascomycetes, which is soluble in alcohol.

The foregoing simple experiments only show (*a*) that the colourless sclerotia possess a direct oxidase system of the laccase type in a region corresponding exactly with that in which the black pigment is present in the sclerotia of the parental strain; and (*b*) that flavones or chromogens are lacking in the colourless sclerotia, or if present, they are inactivated in some way not understood. The experiments do not prove that the pigment in the sclerotia of the parental strain is due to the oxidation of a chromogen or that the aberrant strain has arisen by the loss or inactivation of a factor for chromogen. When, however, the above results are considered along with those of Gortner (32), Keeble and Armstrong (41, 42, 43) and others, which cumulatively form such positive circumstantial evidence, it is somewhat difficult to resist drawing the deduction that the albino strain of *Botrytis cinerea* has indeed arisen from the black parental form by the loss or modification of a factor for chromogen.

## 7. Significance of Colourless Form.

### A. *Discussion of Recorded Mutations.*

The essential facts in the preceding pages may be summarised as follows: The colourless form arose spontaneously without any evident relation to external conditions or stimuli, in a single sclerotium of a culture which on accepted criteria was a pure line. The new form is apparently differentiated from its parent in respect to a single character only, that of colour, and this albinism would appear to be associated with the absence of a factor for chromogen. The change has occurred once only, and has given rise to a form unknown in nature and perfectly constant under all conditions. It would seem possible to place only one interpretation on these facts, that of mutation, and accordingly when I was privileged to exhibit this form to the Linnean Society of London on April 3, 1919, I described it as "an albino mutant of *Botrytis cinerea*."

Previous records of mutation in the fungi fall into two groups. In the first are those changes in genetic constitution which appear to bear a direct and purposive relation to certain conditions which have operated during the development of the organism or that of the preceding generation. As an illustration may be instanced the permanent acquirement of the power to ferment lactose by an organism previously unable to form lactase, this change being brought about by the growth of

the organism or its progenitor upon a lactose medium. The second category of mutations includes those cases in which the genetic change is apparently quite fortuitous, and this group is divisible into two sub-groups. In the first of these the change is associated with certain conditions which interfere with the normal metabolic reactions of the organism. For example, by treating the organism with toxic substances, or by subjecting it to extremes of temperature or desiccation, various indeterminate morphological or physiological changes are induced, such as alterations in spore dimensions or coloration, or growth upon standardised media. In the second sub-group the changes occur spontaneously under, so far as may be judged, perfectly normal conditions of development, and this category would include the colourless form of *Botrytis cinerea*.

Of the purposive mutations the classical case perhaps is that described by Massee (51), in which *Tricothecium roseum* was changed into a virulent parasite by habituating it to a previously immune host. This research to which credence is still largely given was, even for the year 1905, so inexact in all critical details that it may be now be regarded as of merely historical interest. Results worthy of serious consideration are described in the researches of Salmon (66, 67) on the adaptive parasitism of *Erysiphe Graminis* by means of growth upon bridging hosts, or the parallel results obtained with various rust-fungi by Marshall Ward (81, 82), Freeman (28, 29), Johnson (38), Pole Evans (62), and others. In none of this work, however, were the experiments carried out with single-spore strains of the fungi, under the rigidly controlled conditions imperative in such work. Nor was it realised until the more recent researches of Mains (50), Brooks and Cooley (15), Lauritzen (46), and others, how vital are the effects of even minute changes in the light, temperature, age, or humidity relations of the host plants. The findings of Kidd and West (44) on the importance of physiological predetermination in the life of the plant may also have a very material bearing on questions of relative susceptibility or immunity of host plants to fungal attack, and indicate factors which will be extremely difficult to control. Apart, however, from such collateral evidence as this, the value of bridging species in changing the physiological characters of the fungus is not supported by the work of Biffen (12), and has been directly contravened by the researches of Reed (64), and of Stakman and his associates (71–74). After much intensive study of this subject, the latter conclude "... the writers have not been able to detect any mutation nor to induce perceptible evolutionary changes experimentally." With rare exceptions such conclusions are supported by the invariable experience of all the critical and accurate workers in more recent experimental mycology, and there would appear to be little doubt that when a pure line organism is examined by a careful investigator under rigidly controlled conditions, heritable alterations bearing a purposive relation to educative treatment are absent.

The recorded instances of non-purposive genetic changes in organisms induced by

exposure to unusual conditions are very few in number. In 1908 Arcichovskij (2) described a form of *Aspergillus niger* with yellow-brown spores which arose in a culture of the black form growing in Raulin's fluid to which 0·0001 per cent. of zinc sulphate had been added. Four years later Schiemann (69) obtained various mutants of *Aspergillus niger* by treatment of cultures with potassium bichromate, exposure to high temperatures, and so forth.

In the same year Waterman (83) caused *Aspergillus niger* to mutate by treatment with 2 per cent. galactose, rhamnose, or glucose, 1 per cent. boric acid, *p*-oxybenzoic acid, or dichloracrylic acid. A fungus which the author speaks of as *Penicillium glaucum* was also caused to throw mutants by treatment with any of the following substances:—*p*-oxybenzoic acid, salicylic acid, trichloracrylic acid, tetrachlor-propionamid, pentachlor-propionamid, and pyrocatechetic acid (pyrocatechuic?).

In all the above cases the new forms described differed from the parental fungi in many morphological and physiological characters, the most obvious being that of the colour of the spores. Furthermore, the production of the mutant forms was not a constant and specific reaction to certain definite chemical stimuli, but was merely a fortuitous circumstance resulting from the generally unfavourable conditions to which the organisms had been subjected. These were of such diverse nature that one might expect that almost any substance deleterious to growth or interfering with the metabolic processes in any way would produce similar results. Also not every culture subjected to these conditions gave uniform results over the entire growth, but only occasional cultures, and in these cultures only sporadic individuals, conidiophores or spores. Moreover, as Schiemann (69) points out, mutations were also produced when the fungi were developing under favourable conditions.

Such results as these would have a somewhat indifferent reception if obtained in a physical or chemical laboratory. Their anomalous nature is stated clearly in Schiemann's paper as follows: "Die konstanten Farbmutationen sind beide auf mit $K_2Cr_2O_7$ versetzten Nährböden aufgetreten. Von einer spezifischen Wirkung des $K_2Cr_2O_7$ zu sprechen, ist trotzdem nicht möglich. Schon der Umstand, dass der Farbumschlag unter über 100 Kulturen mit $K_2Cr_2O_7$ nur zweimal stattfand, hier wiederum nicht die ganze Decke betraf, sondern unter vielen hunderten von Köpfchen nur einzelne, spricht dagegen. Ferner hat die *Proteus*-Mutante dieselbe Richtung eingeschlagen; diese aber ist aus Hitzekulturen, einmal mit, einmal ohne Zusatz von $K_2Cr_2O_7$ hervorgegangen. Endlich sind nach braun abändernde Kulturen des *Aspergillus niger* auch sonst gelegentlich beobachtet worden—wenn auch selten—ohne dass das Chromal zur Verwendung kam."

In a discussion of the causes of genetic variation, Bateson (7) has written: "The state of knowledge of this whole subject is . . . most unsatisfactory, chiefly for the reason that in none of the cases which are alleged to show a positive result have two observers been over the same ground, or as yet confirmed each other . . . I do not know a single case which has been established and confirmed in such a way that we

could with confidence expect to witness the alleged phenomena if we were to repeat the experiment."

In the work of Arcichovskij (2), Schiemann (69), and Waterman (83), mutant forms were produced so easily and with such a diversity of stimuli and simplicity of technique, that it seemed probable that a repetition of their experiments might produce the necessary confirmatory evidence. Accordingly single-spore cultures of two strains of *Aspergillus niger* and of two strains of *Penicillium Italicum* were prepared. (Waterman used "*Penicillium glaucum*," but as Thom (77) points out, this "name as used at present seems to be applied collectively to the common green forms, which under examination are quickly found to be not one but several species.") With these fungi, the experiments of Schiemann (69) and Arcichovskij (2), and certain of those of Waterman (83) (galactose, glucose, rhamnose, boric acid, and salicylic acid) were carefully repeated. Although modifications in the colour of the spores, general morphological facies, and physiological reactions were observed, these affected the whole growth operated upon and to an equal degree; but the reproductive bodies of the modified fungi, when returned to the original conditions, gave in every case the original result. These changes were phenotypic and not genotypic, and throughout the whole series of experiments not a single mutation occurred.

More recently, experiments have been carried out with another strain of *Aspergillus niger* isolated from the air, and with a strain of *Penicillium* (*roseum ?*) from the soil, both these having been in culture in the laboratory for several months and repeatedly subcultured, although not derived from single spores. The fungi were subjected to rather gross treatment with various dilutions of toxic substances, such as silver nitrate, copper sulphate, mercury bichloride, lead nitrate, potassium acid phosphate, sodium chloride, and so forth, and then platings were made from these tubes. The resulting growths were perfectly normal.

In all work on the induction of genotypic alterations in micro-organisms, it is extremely difficult, even when the most scrupulous attention is given to every detail, to be quite certain that one has under control each of the almost infinite number of factors which may influence the results. Mycologists are only now beginning to learn how extremely complex and difficult to isolate such factors may be, and the researches of, for example, Duggar (25) and his collaborators on the physiology of certain fungi grown in culture media, or those of Coons (21) on the factors involved in the growth and pycnidium production of *Plenodomus fuscomaculans*, indicate the complexities underlying apparently the most simple physiological phenomena. Not the least interesting of the many experimental studies throwing light on the bionomics of *Aspergillus niger* are those of Javillier (34) and of Steinberg (75), in which it is shown how even such an unsuspected factor as the different quantities of zinc present in the glass of flasks, test-tubes, and other apparatus may vitiate comparative cultural studies of this fungus.

If, in the work of Arcichovskij (2), Schiemann (69), and Waterman (83),

accidental contamination did not occur, one is almost forced to conclude that some accessory factor of an unrecognised nature must have been operative in their experiments; a factor absent in my own repetition of their work and absent from the studies of every other mycologist who has carried out critical experimental investigations on these fungi.

Furthermore, as has been pointed out, these investigators emphasise the fact that their mutations are not of the nature of specific reactions to definite chemical or physical stimuli, but only the results of a generalised interference with the normal course of life of the organism. Thus SCHIEMANN (69) says: " . . . so zeigt sich als ein Gemeinsames, dass in allen Fällen, wo derartige Farbänderungen auftraten, Störungen der normalen Lebens-verhältnisse vorlagen. Es ist deshalb der Schluss berechtigt, dass die Mutation in den beobachteten Fällen durch einen starken Reiz ausgelöst wurde." But the experience of other investigators demonstrates the extreme genetic constancy of species of these fungi even when placed for many generations under the most unfavourable conditions. Numerous workers have shown that the morphological facies and physiological activities of fungi may within certain limits be changed at will in the single generation, but in no case, with the above exceptions, have these phenotypic modifications had any effect whatever on the genotypic constitution of the organisms.

### B. *Alternative Interpretations.*

If the recorded cases of mutation in the fungi need not be accepted—and for my own part I do not feel compelled to accept them—how then is one to regard the colourless strain of *Botrytis cinerea*, which has been described in the present paper? Its fortuitous origin as a single sclerotium in a perfectly normal single-spore strain, whose development since its commencement is known, has already been emphasised. Under no conditions whatever, so far as is known, may the black sclerotial pigment of the parent strain, or the albinism of its derivative, be modified, and although other characters may show plasticity, no instance of a permanent alteration has been observed. Does the phenotypic constancy of the parental strain under constant conditions necessarily imply genotypic purity?

In the preceding account the parental strain has been spoken of as a "pure line," and on all accepted criteria this description is merited, for not only is it a single spore strain of an asexual organism, but it has been reproduced for most of its generations either by spores from single conidiophores or from single germinated sclerotia, both of which take origin in a single cell. In his work on 'Evolution by means of Hybridisation,' LOTSY (48) rightly insists that: "*Certainty of purity is a* conditio sine qua non *to obtain proof of the existence of mutation in living beings, just as chemical purity is a* conditio sine qua non *to obtain proof of the existence of mutability of the elements.*" Is there certainty of purity in the parental strain of *Botrytis cinerea* which gave rise to the colourless form? Further, is the fulfilment of such a

criterion possible in the fungi (or bacteria), and particularly in those forms for which mutation is claimed?

The only positive contribution to this subject is the work of Burgeff (18), published in 1914, in which the results of the crossing of various forms of *Phycomyces nitens* are described. This fungus possesses a cœnocytic mycelium, containing numerous scattered nuclei. In the asexual form of reproduction spores are delimited within sporangial heads into which there have passed an indefinite number of nuclei. The multinucleate spores germinate and reproduce the cœnocytic mycelium. If, therefore, the original hyphæ are genetically impure, this condition will be maintained in all succeeding generations, for the sporangiospores merely reproduce the genetic condition of the hyphæ which give rise to the sporangia. Opportunity for genetic contamination occurs at sexual reproduction, for this process is merely a fusion of two multinucleate gametes to form a multinucleate zygospore, which on germination gives rise to a cœnocytic mycelium containing nuclei of both parental strains. There will be an equal chance for both types of nuclei to pass into the sporangia and be included in the multinucleate spores. A single-spore strain may thus be heterocaryotic. If now this form at sexual maturity fuses with a third form, and so on, the genotype of any particular isolation may be extremely complex. As, moreover, the sporangiospores are delimited and the walls of the zygogametes laid down without any apparent regard either to the condition, the number, or the position of the nuclei they separate, there is no absolute surety that two single-spore strains derived from an original single-spore strain will have the same genotype.

In such *Mucorineæ* it is, therefore, totally impossible to comply with the criterion of specific purity which is imperative in this issue. There is an almost equal lack of certainty of specific purity in the genera *Aspergillus* and *Penicillium*. The mycelium of these fungi is multicellular and each cell contains many nuclei. On asexual reproduction a vegetative cell gives rise to a conidiophore, and from this conidia are abstricted. The hyphal mother cell is multinucleate and these nuclei, by division and wandering, finally pass into the reproductive cells, these, in many species of the fungi, containing an indefinite number of nuclei. In other species the conidia generally contain but a single nucleus, but almost any preparation of such conidia shows multinucleate individuals.

A single-spore strain thus reproduces the genotype of its parent, and if the latter contains nuclei of different constitutions, this heterocaryotic condition will be maintained in the progeny.

Now, in the case of *Phycomyces*, genetic contamination occurs in the sexual process, for many of these forms are allogamous. There are, however, no species of *Aspergillus* or *Penicillium* known in which sexuality is necessarily allogamous, although there is no certainty that this type of reproduction may not occur, at all events, at intervals. We still know too little of what takes place in the sexual processes of either of these genera to make their discussion profitable; but given a

union of two parental mycelia only once in many thousands of generations, the mycelia of the offspring can afford no certainty of specific purity.

Moreover in any culture of a single-spore strain of either fungus it is not usually difficult to find two hyphæ which have come into contact, and the walls separating them have been auto-digested, giving cytoplasmic continuity. In a mixed growth of a number of strains these somatic fusions are also to be found, and although one is not able to state that fusions occur between hyphal cells of different strains, neither may the rare possibility of this be denied. If now from such a contaminated cell in *Aspergillus* or *Penicillium* a conidiophore were to arise, or either or both the sexual organs, the offspring of the reproductive bodies would be genetically impure, and asexual reproduction would only maintain this heterocaryotic or heterozygotic condition.

Thus even accepting the evidence of Arcichovskij (2), Schiemann (69), and Waterman (83), it by no means follows that the permanent changes they observed are necessarily mutations, for they may be interpreted in terms of genetic contamination, and in no case may these organisms comply with that criterion of "certainty of specific purity which is a *conditio sine qua non* to obtain proof of the existence of mutation." Although in this account very little reference has been made to bacterial mutation, there can be little doubt that the possibility of the rare occurrence of "somatic" fusions in these organisms places them out of consideration as subject organisms for the experimental induction of mutations.

Thus, according to the principle of parsimony, all so-called mutations in these groups of organisms may perhaps more wisely be interpreted in terms of the splitting of an originally impure genetic constitution or of gametic or somatic segregation from heterozygotes.

These considerations applied to the fungus *Botrytis cinerea* throw an entirely new light upon the value to be attached to the colourless strain. *Botrytis cinerea* possesses a multicellular mycelium, each cell of which contains many nuclei. The conidiophores are multinucleate and a small but indefinite number of nuclei pass into each conidium. Throughout the whole of the vegetative and reproductive mycelium the septa are laid down by a diaphragm-like growth from the hyphal walls, irrespective of the number or condition of the nuclei thus separated. Each conidium, therefore, merely reproduces the genetic constitution of the original cell of the mycelium in which its conidiophore arose. Many thousands of such asexual generations would, therefore, not alter the genetic constitution of the organism, and there is no sexual process. The possibility of genetic contamination is brought about by the occurrence of hyphal anastomoses. In the extremely rare chance of a fertile conidiophore arising from a cell contaminated by the nuclei or cytoplasm of a genotypically different individual lies, I believe, the explanation of the colourless form of *Botrytis* described. While this possibility exists, it is more consonant with the principles of scientific methodology to accept this interpretation than to formulate a mutational hypothesis of the origin of the aberrant strain.

Before concluding, a few words may perhaps be said concerning the *rôle* of the nucleus in the heredity of multinucleate fungi. The conidia of *Aspergillus*, *Penicillium* or *Botrytis* contain a variable number of nuclei, and yet if conidia are plated out the colonies which develop are essentially similar in physiological reactions and morphological properties for each particular species. It follows, therefore, that the exact number of nuclei which chance to be enclosed in each reproductive body has no bearing on the absence or presence of particular characters in the progeny, but is a purely fortuitous occurrence. What is vital, is, as Burgeff (18) has shown, not the number of nuclei in each cell but the kind of nuclei present. Although the evidence on this point is meagre there would yet seem no reason to doubt that the structure of the nucleus in fungi is in general fundamentally similar to that in other organisms.

Following Morgan (55) the chromosomes might thus be visualised as linear series of loci, like threaded beads. At each locus exist specific factors, making up a reaction system the elements of which bear a more or less specific relationship to one another. Thus, eight loci, representing the factor for black sclerotial pigment in the chromosomes of as many identical nuclei, would therefore have neither greater nor less visible expression than one locus in the chromosome of a single nucleus, or than 15 loci in the chromosomes of 15 nuclei in a single reproductive body. Increased number of nuclei does not signify a cumulative value for any particular character. But if owing to somatic fusion a hyphal cell containing eight nuclei with loci for black pigment became possessed of one or more contaminating nuclei from a genotypically different individual, and if this particular cell gave rise to a sclerotium, the characters of this sclerotium would be determined by the reacting systems of loci present in the two kinds of nuclei, the visible characters being the expression of the resultant of their interaction. The contaminating nuclei might find themselves in a cytoplasmic environment so foreign to their normal metabolism as to remain inert, or the locus for colour might be counterbalanced or cancelled by the locus for black in any one of the eight original nuclei, in which case the seven remaining nuclei would reproduce the original colour. In both cases the sclerotial characters would be wholly normal. On the other hand the loci in the invading nuclei might function as lethal factors, such as Morgan (54) has shown to exist in the case of *Drosophila*, and development might be partially or entirely inhibited.

Again, the factors in the one system might react in many ways with those of the other system changing the physiological and morphological characters of the adult form so that apparently spontaneous variation in several independent directions might result, producing "mutants" such as those described by Arcichovskij, Schiemann and Waterman. The majority of the factors of the contaminating nuclei might remain inert, but one or more might react with certain factors in the original nuclei resulting in the inhibition of one or more of the latter. If, for example, the factor for chromogen were thus inhibited or destroyed the adult would be an albino similar to the new form of *Botrytis cinerea* described. If there are several factors for

colour, as would appear probable in the present instance, that associated with sclerotial pigmentation only might be inhibited, leaving unaltered the factors governing the pigmentation of spores or haptera.

On the other hand, it is not improbable that the contamination may be purely cytoplasmic, and the progeny of such a cell will therefore still be homocaryotic. The foreign cytoplasm, however, would create a new environment through which the nuclei would function, and this might so influence the development of the organisms as to alter materially the outward expression of one or more characters in the adult form. As the added cytoplasm would now be an integral portion of the organism, this changed expression would be permanent and heritable. It is thus possible for apparent "mutations" to occur in the fungi without any change in nuclear character. So far as the nucleus is concerned, regarding this structure as an entity distinct from the cytoplasm, the change would merely be phenotypic; but for the complete organism the change would be genotypic. It is perhaps in such purely cytoplasmic contaminations that apparent mutations in those organisms having uninucleate cells may be sought.

In the case under discussion, nuclear and cytoplasmic contamination must have occurred prior to the isolation of the single-spore strain, and yet the colourless strain arose only in the third tube-generation subsequent to this. If, however, one assumes a contaminated original strain, it is not inconceivable that in the fortuitous septation of the hyphæ, certain nuclei and cytoplasmic matter were separated, such that in all things save chromogen formation the resultant resembled the parental strain. The expression of this resultant would be an albino strain such as that described. Such a concomitance of conditions might well appear almost inconceivable, but the appearance of an albinistic strain of the fungus is equally almost inconceivable, for in the history of the genus as we know it it has occurred but once.

It would seem to me probable that the same concept which Bateson (8) and Lotsy (48) have formulated for the higher organisms may hold true for both fungi and bacteria, but here the recombination of genes may not only take place by sexual fusions, but also by direct "contamination" of one species by the nuclei and cytoplasm of another in vegetative anastomoses. New reacting systems of chromosomal loci impinge upon each other, and the resultant is the new form. This at first may be heterocaryotic, but if true sexual processes occur later, the offspring become homocaryotic but heterozygous. Such a hypothesis fits the facts and explains them as I think no other present hypothesis does.

It is not the possibility of mutation in the fungi (and bacteria) that is here denied, but rather the compulsion to accept that interpretation of the evidence as it now stands. As the facts present themselves to me, avenues of interpretation other than that of mutation are still widely open.

## BIBLIOGRAPHY.

(1) Abderhalden, E., and Guggenheim. 'Zeitsch. Physiol. Chem.,' vol. 54 (1907), and vol. 57 (1908).
(2) Arcichovskij, V. *Autoreferat* in 'Centr. f. Bakt.,' II., vol. 21 (1908).
(3) Bach, A. 'Rev. Deut. Chem. Ges.,' vol. 41 (1908), and vol. 42 (1909).
(4) *Idem.* 'Biochem. Zeits.' vol. 60 (1914).
(5) Bary, A. de. 'Morph. u. Phys. der Pilze,' etc. (1884).
(6) *Idem.* 'Bot. Zeit.,' vol. 44 (1886).
(7) Bateson, W. 'Problems of Genetics' (1913).
(8) *Idem.* 'Rept. Brit. Ass.' (1914).
(9) Beauverie, J., and Guilliermond, A. 'Compt. Rend.,' vol. 128 (1899), and vol. 133 (1901).
(10) *Idem.* 'Centr. f. Bakt.,' II, vol. 10 (1903).
(11) Bertrand, G. 'Compt. Rend.,' vols. 122 and 123 (1896).
(12) Biffen, R. H. 'Journ. Agr. Sci.,' vol. 4 (1912).
(13) Blackman, V. H., and Welsford, E. J. 'Anns. Bot.,' vol. 30 (1916).
(14) Bourquelot, E. M., and Bertrand, G. 'Bull. Soc. Mycol. de France,' vol. 12 (1896).
(15) Brooks, C., and Cooley, J. S. 'Journ. Agr. Res.' vol. 8 (1917).
(16) Brown, W. 'Anns. Bot.,' vol. 29 (1915), vol. 30 (1916).
(17) *Idem.* 'New Phyt.,' vol. 16 (1917).
(18) Burgeff, H. 'Flora,' vol. 7 (1914).
(19) Burri, R. 'Das Tuscheverfahren' (1909).
(20) Chodat, R., and Staub. 'Arch. Sci. Phys. Nat.,' vol. 23 and 24 (1907).
(21) Coons, G. H. 'Journ. Agr. Res.,' vol. 5 (1916).
(22) Crabill, C. H., and Reed, H. S. 'Biochem. Bull.,' vol. 4 (1915).
(23) Dox, A. W. U.S. Dept. Agr., 'Bur. An. Ind., Bull.,' No. 120 (1910).
(24) Drechsler, C. 'Bot. Gaz.,' vol. 67 (1919).
(25) Duggar, B. M., and Collaborators. 'Anns. Mo. Bot. Gard.' (1916–1919).
(26) Elliott, J. A. 'Amer. Journ. Bot.,' vol. 4 (1917).
(27) Farneti, R. 'Atti R. Ist. Bot. Univ. Pavia.,' vol. 7 (1902).
(28) Freeman, E. M. 'Anns. Bot.,' vol. 5 (1902).
(29) Freeman, E. M., and Johnson, E. C. 'U.S. Dept. Agr. Bur. Pl. Ind. Bull.,' No. 216 (1911).
(30) Gaumann, E. 'Anns. Mycol.,' vol. 16 (1918).
(31) *Idem.* 'Beih. z. Bot. Centr.,' vol. 35 (1918).
(32) Gortner, R. A. 'Amer. Nat.,' vol. 45 (1911).
(33) Istvanffi, Gy. de. 'Anns. de l'Inst. Cent. Ampel. R. Hongr.,' vol. 3 (1905).
(34) Javillier. 'Compt. Rend.,' vol. 158 (1914).
(35) Jennings, F. 'Genetics,' vol. 1 (1916), and vol. 2 (1917).

(36) Johannsen, W. 'Über Erblichkeit in Populationen und in reinen Linien' (1903).
(37) *Idem.* 'Elemente der exakten Erblichkeitslehre' (1909).
(38) Johnson, E. C. 'U.S. Dept. Agr. Bur. Pl. Ind. Bull.,' No. 224 (1911).
(39) Jones, W. N. 'Proc. Roy. Soc.,' B, vol. 85 (1912).
(40) Kastle, J. H. 'Hyg. Lab. U.S. Pub. Health Bull.,' No. 26 (1906).
(41) Keeble, F., and Armstrong, E. F. 'Proc. Roy. Soc.,' B, vol. 85 (1912).
(42) *Idem.* 'Journ. Genetics,' vol. 2 (1912).
(43) Keeble, F., Armstrong, E. F., and Jones, W. N. 'Proc. Roy. Soc.,' B, vol. 86 (1913), and vol. 87 (1913).
(44) Kidd, F., and West, C. 'Anns. Applied Biol.,' vol. 5 (1918–19).
(45) Kissling, M. 'Hedwigia,' vol. 28 (1889).
(46) Lauritzen, J. I. 'Phytopath,' vol. 9 (1919).
(47) Lind, J. 'Danish Fungi in the Herbarium of E. Rostrup' (1913).
(48) Lotsy, J. P. 'Evolution by Means of Hybridization' (1916).
(49) Lutz, M. L. 'Bull. Soc. Mycol. France,' vol. 28 (1912).
(50) Mains, E. B. 'Amer. Journ. Bot.,' vol. 4 (1917).
(51) Massee, G. 'Phil. Trans.,' B, vol. 197 (1905).
(52) Möbius, M. 'Ber. D. Bot. Ges.,' vol. 18 (1900).
(53) Moreau, F. 'Bull. Soc. Myc. France,' vol. 33 (1917).
(54) Morgan, T. H. 'Journ. Expt. Zool.,' vol. 17 (1914).
(55) Morgan, T. H., Sturtevant, A. H., Muller, H. J., and Bridges, C. B. 'The Mechanism of Mendelian Heredity' (1915).
(56) Munn, M. T. 'Geneva Agr. Expt. Sta. N.Y. Bull.,' No. 437 (1917).
(57) Mutto, E., and Pollacci, G. 'Rendic. R. Accad. Lincei,' vol. 26 (1917).
(58) Nordhausen, M. 'Jahrb. f. Wiss. Bot.,' vol. 33 (1899).
(59) Peltier, G. L. 'Anns. Mo. Bot. Gard.,' vol. 23 (1912).
(60) Pethybridge, G. H. 'Journ. Dept. Agr., Ireland,' vol. 11 (1911), and vol. 16 (1916).
(61) Pirotta. 'Nuovo Giorn. Bot. Ital.,' vol. 13 (1881).
(62) Pole Evans, J. B. 'Journ. Agr. Sci.,' vol. 4 (1911).
(63) Pringsheim, H. 'Zeitsch. Physiol. Chem.,' vol. 62 (1909).
(64) Reed, G. M. 'Memrs. Brooklyn Bot. Gard.,' vol. 1 (1918).
(65) Reidemeister, W. 'Anns. Mycol.,' vol. 7 (1909).
(66) Salmon, E. S. 'Anns. Mycol.,' vol. 2 (1904).
(67) *Idem.* 'Beih. z. Bot. Centr.,' vol. 14 (1903).
(68) Seaver, S., and Horne, W. T. 'Memrs. Torr. Bot. Club,' vol. 17 (1918).
(69) Schiemann, E. 'Zeit. f. Indukt. Abstamm.,' vol. 8 (1912).
(70) Smith, R. E. 'Bot. Gaz.,' vol. 29 (1900), and vol. 33 (1903).
(71) Stakman, E. C. 'Minn. Agr. Expt. Sta. Bull.,' No. 138 (1914).
(72) Stakman, E. C., and Jenson, L. 'Journ. Agr. Res.,' vol. 5 (1915).

(73) Stakman, E. C., and Piemeisel, F. J. 'Journ. Agr. Res.,' vol. 10 (1917).
(74) Stakman, E. C., Piemeisel, F. J., and Levine, M. 'Journ. Agr. Res.,' vol. 15 (1918).
(75) Steinberg, R. A. 'Memrs. Torr. Bot. Club,' vol. 17 (1918).
(76) Stevens, F. L., and Hall, J. G. 'Bot. Gaz.,' vol. 48 (1909).
(77) Thom, C. 'U.S. Bur. An. Ind. Bull.,' No. 118 (1910).
(78) *Idem.* 'Journ. Agr. Res.,' vol. 7 (1916).
(79) Viala, P. 'Maladies de la Vigne' (1893).
(80) Ward, H. M. 'Anns. Bot.,' vol. 2 (1888).
(81) *Idem.* 'Anns. Bot.,' vol. 5 (1901).
(82) *Idem.* 'Anns. Mycol.,' vol. 1 (1903).
(83) Waterman, H. J. 'Kon. Akad. Wetensch. Amsterdam,' vol. 15 (1912).
(84) Welsford, E. J. 'Anns Bot.,' vol. 30 (1916).
(85) Wheldale, M. 'The Anthocyanin Pigments of Plants' (1916).
(86) Wheldale, Onslow, M. 'Biochem. Journ.,' vol. 13 (1919).
(87) Zeller, S. M. 'Anns. Mo. Bot. Gard.,' vol. 3 (1916).
(88) Zellner. 'Die Chemie der Höheren Pilze' (1907).
(89) Zopf, W. 'Die Pilze' (1890).

## DESCRIPTION OF PLATE.

Fig. 1.—Petri dish culture in which the normal form and its colourless derivative have been grown. Pigmented haptera of the albino strain may be seen at the periphery of the culture in contact with the glass. × $\frac{2}{3}$.

Fig. 2.—Vertical section of colourless sclerotium. × 250.

Fig. 3.—Vertical section of parental sclerotium. × 250.

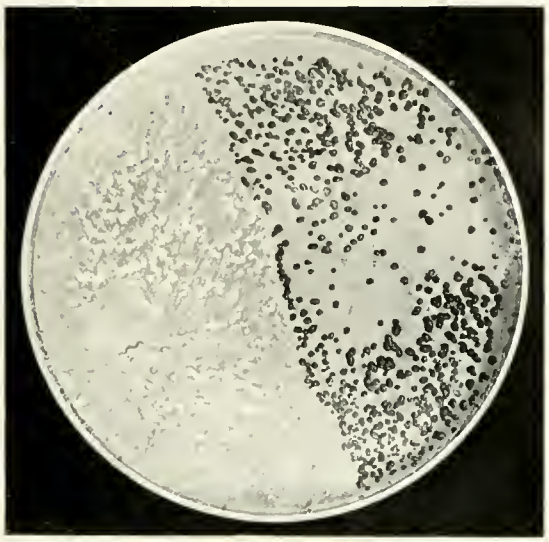


FIG. 1.

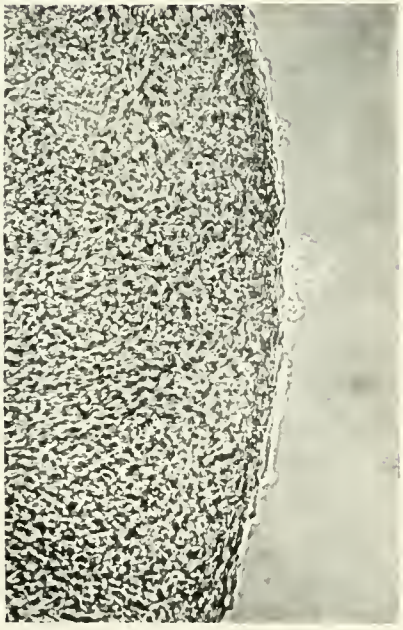


FIG. 2.

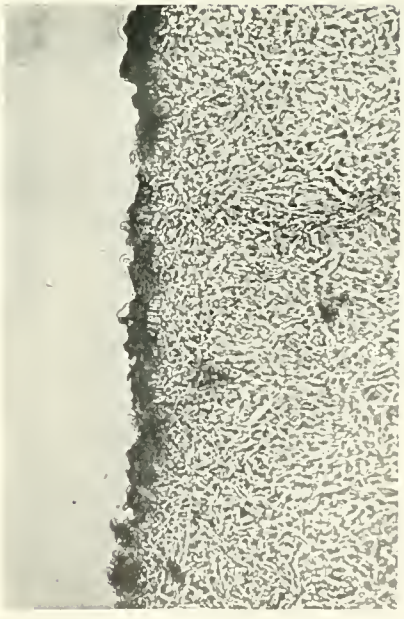


FIG. 3.

# IX. *On the Mathematical Foundations of Theoretical Statistics.*

*By* R. A. Fisher, M.A., *Fellow of Gonville and Caius College, Cambridge, Chief Statistician, Rothamsted Experimental Station, Harpenden.*

*Communicated by* Dr. E. J. Russell, F.R.S.

Received June 25,—Read November 17, 1921.

## Contents.

| Section | | Page |
|---|---|---|
| 1. | The Neglect of Theoretical Statistics | 310 |
| 2. | The Purpose of Statistical Methods | 311 |
| 3. | The Problems of Statistics | 313 |
| 4. | Criteria of Estimation | 316 |
| 5. | Examples of the Use of Criterion of Consistency | 317 |
| 6. | Formal Solution of Problems of Estimation | 323 |
| 7. | Satisfaction of the Criterion of Sufficiency | 330 |
| 8. | The Efficiency of the Method of Moments in Fitting Curves of the Pearsonian Type III | 332 |
| 9. | Location and Scaling of Frequency Curves in general | 338 |
| 10. | The Efficiency of the Method of Moments in Fitting Pearsonian Curves | 342 |
| 11. | The Reason for the Efficiency of the Method of Moments in a Small Region surrounding the Normal Curve | 355 |
| 12. | Discontinuous Distributions | 356 |
| | (1) The Poisson Series | 359 |
| | (2) Grouped Normal Data | 359 |
| | (3) Distribution of Observations in a Dilution Series | 363 |
| 13. | Summary | 366 |

## Definitions.

*Centre of Location.*—That abscissa of a frequency curve for which the sampling errors of optimum location are uncorrelated with those of optimum scaling. (9.)

*Consistency.*—A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter. (4.)

*Distribution.*—Problems of distribution are those in which it is required to calculate the distribution of one, or the simultaneous distribution of a number, of functions of quantities distributed in a known manner. (3.)

*Efficiency.*—The efficiency of a statistic is the ratio (usually expressed as a percentage) which its intrinsic accuracy bears to that of the most efficient statistic possible. It

expresses the proportion of the total available relevant information of which that statistic makes use. (4 and 10.)

*Efficiency* (*Criterion*).—The criterion of efficiency is satisfied by those statistics which, when derived from large samples, tend to a normal distribution with the least possible standard deviation. (4.)

*Estimation.*—Problems of estimation are those in which it is required to estimate the value of one or more of the population parameters from a random sample of the population. (3.)

*Intrinsic Accuracy.*—The intrinsic accuracy of an error curve is the weight in large samples, divided by the number in the sample, of that statistic of location which satisfies the criterion of sufficiency. (9.)

*Isostatistical Regions.*—If each sample be represented in a generalized space of which the observations are the co-ordinates, then any region throughout which any set of statistics have identical values is termed an isostatistical region.

*Likelihood.*—The likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed.

*Location.*—The location of a frequency distribution of known form and scale is the process of estimation of its position with respect to each of the several variates. (8.)

*Optimum.*—The optimum value of any parameter (or set of parameters) is that value (or set of values) of which the likelihood is greatest. (6.)

*Scaling.*—The scaling of a frequency distribution of known form is the process of estimation of the magnitudes of the deviations of each of the several variates. (8.)

*Specification.*—Problems of specification are those in which it is required to specify the mathematical form of the distribution of the hypothetical population from which a sample is to be regarded as drawn. (3.)

*Sufficiency.*—A statistic satisfies the criterion of sufficiency when no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter to be estimated. (4.)

*Validity.*—The region of validity of a statistic is the region comprised within its contour of zero efficiency. (10.)

## 1. The Neglect of Theoretical Statistics.

Several reasons have contributed to the prolonged neglect into which the study of statistics, in its theoretical aspects, has fallen. In spite of the immense amount of fruitful labour which has been expended in its practical applications, the basic principles of this organ of science are still in a state of obscurity, and it cannot be denied that, during the recent rapid development of practical methods, fundamental problems have been ignored and fundamental paradoxes left unresolved. This anomalous state of statistical science is strikingly exemplified by a recent paper (1) entitled " The Funda-

mental Problem of Practical Statistics," in which one of the most eminent of modern statisticians presents what purports to be a general proof of BAYES' postulate, a proof which, in the opinion of a second statistician of equal eminence, "seems to rest upon a very peculiar—not to say hardly supposable—relation." (2.)

Leaving aside the specific question here cited, to which we shall recur, the obscurity which envelops the theoretical bases of statistical methods may perhaps be ascribed to two considerations. In the first place, it appears to be widely thought, or rather felt, that in a subject in which all results are liable to greater or smaller errors, precise definition of ideas or concepts is, if not impossible, at least not a practical necessity. In the second place, it has happened that in statistics a purely verbal confusion has hindered the distinct formulation of statistical problems; for it is customary to apply the same name, *mean, standard deviation, correlation coefficient*, etc., both to the true value which we should like to know, but can only estimate, and to the particular value at which we happen to arrive by our methods of estimation; so also in applying the term probable error, writers sometimes would appear to suggest that the former quantity, and not merely the latter, is subject to error.

It is this last confusion, in the writer's opinion, more than any other, which has led to the survival to the present day of the fundamental paradox of inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts. The criticisms of BOOLE, VENN, and CHRYSTAL have done something towards banishing the method, at least from the elementary text-books of Algebra; but though we may agree wholly with CHRYSTAL that inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply committed itself), there yet remains the feeling that such a mistake would not have captivated the minds of LAPLACE and POISSON if there had been nothing in it but error.

## 2. THE PURPOSE OF STATISTICAL METHODS.

In order to arrive at a distinct formulation of statistical problems, it is necessary to define the task which the statistician sets himself: briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample. The law of distribution of this hypothetical population is specified by relatively few parameters, which are sufficient to describe it exhaustively in respect of all qualities under discussion. Any information given by the sample, which is of use in estimating the values of these parameters, is relevant information. Since the number of independent facts supplied in

the data is usually far greater than the number of facts sought, much of the information supplied by any actual sample is irrelevant. It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data.

When we speak of the *probability* of a certain object fulfilling a certain condition, we imagine all such objects to be divided into two classes, according as they do or do not fulfil the condition. This is the only characteristic in them of which we take cognisance. For this reason probability is the most elementary of statistical concepts. It is a parameter which specifies a simple dichotomy in an infinite hypothetical population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit. For example, when we say that the probability of throwing a five with a die is one-sixth, we must not be taken to mean that of any six throws with that die one and one only will necessarily be a five; or that of any six million throws, exactly one million will be fives; but that of a hypothetical population of an infinite number of throws, with the die in its original condition, exactly one-sixth will be fives. Our statement will not then contain any false assumption about the actual die, as that it will not wear out with continued use, or any notion of approximation, as in estimating the probability from a finite sample, although this notion may be logically developed once the meaning of probability is apprehended.

The concept of a *discontinuous frequency distribution* is merely an extension of that of a simple dichotomy, for though the number of classes into which the population is divided may be infinite, yet the frequency in each class bears a finite ratio to that of the whole population. In *frequency curves*, however, a second infinity is introduced. No finite sample has a frequency curve: a finite sample may be represented by a histogram, or by a frequency polygon, which to the eye more and more resembles a curve, as the size of the sample is increased. To reach a true curve, not only would an infinite number of individuals have to be placed in each class, but the number of classes (arrays) into which the population is divided must be made infinite. Consequently, it should be clear that the concept of a frequency curve includes that of a hypothetical infinite population, distributed according to a mathematical law, represented by the curve. This law is specified by assigning to each element of the abscissa the corresponding element of probability. Thus, in the case of the normal distribution, the probability of an observation falling in the range $dx$, is

$$\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx,$$

in which expression $x$ is the value of the variate, while $m$, the mean, and $\sigma$, the standard deviation, are the two parameters by which the hypothetical population is specified. If a sample of $n$ be taken from such a population, the data comprise $n$ independent facts. The statistical process of the reduction of these data is designed to extract from them all relevant information respecting the values of $m$ and $\sigma$, and to reject all other information as irrelevant.

It should be noted that there is no falsehood in interpreting any set of independent measurements as a random sample from an infinite population ; for any such set of numbers are a random sample from the totality of numbers produced by the same matrix of causal conditions : the hypothetical population which we are studying is an aspect of the totality of the effects of these conditions, of whatever nature they may be. The postulate of randomness thus resolves itself into the question, " Of what population is this a random sample ? " which must frequently be asked by every practical statistician.

It will be seen from the above examples that the process of the reduction of data is, even in the simplest cases, performed by interpreting the available observations as a sample from a hypothetical infinite population ; this is *a fortiori* the case when we have more than one variate, as when we are seeking the values of coefficients of correlation. There is one point, however, which may be briefly mentioned here in advance, as it has been the cause of some confusion. In the example of the frequency curve mentioned above, we took it for granted that the values of both the mean and the standard deviation of the population were relevant to the inquiry. This is often the case, but it sometimes happens that only one of these quantities, for example the standard deviation, is required for discussion. In the same way an infinite normal population of two correlated variates will usually require five parameters for its specification, the two means, the two standard deviations, and the correlation ; of these often only the correlation is required, or if not alone of interest, it is discussed without reference to the other four quantities. In such cases an alteration has been made in what is, and what is not, relevant, and it is not surprising that certain small corrections should appear, or not, according as the other parameters of the hypothetical surface are or are not deemed relevant. Even more clearly is this discrepancy shown when, as in the treatment of such fourfold tables as exhibit the recovery from smallpox of vaccinated and unvaccinated patients, the method of one school of statisticians treats the proportion of vaccinated as relevant, while others dismiss it as irrelevant to the inquiry. (3.)

## 3. The Problems of Statistics.

The problems which arise in reduction of data may be conveniently divided into three types :—

(1) Problems of Specification. These arise in the choice of the mathematical form of the population.

(2) Problems of Estimation. These involve the choice of methods of calculating from a sample statistical derivates, or as we shall call them statistics, which are designed to estimate the values of the parameters of the hypothetical population.

(3) Problems of Distribution. These include discussions of the distribution of statistics derived from samples, or in general any functions of quantities whose distribution is known.

It will be clear that when we know (1) what parameters are required to specify the

population from which the sample is drawn, (2) how best to calculate from the sample estimates of these parameters, and (3) the exact form of the distribution, in different samples, of our derived statistics, then the theoretical aspect of the treatment of any particular body of data has been completely elucidated.

As regards problems of specification, these are entirely a matter for the practical statistician, for those cases where the qualitative nature of the hypothetical population is known do not involve any problems of this type. In other cases we may know by experience what forms are likely to be suitable, and the adequacy of our choice may be tested *a posteriori*. We must confine ourselves to those forms which we know how to handle, or for which any tables which may be necessary have been constructed. More or less elaborate forms will be suitable according to the volume of the data. Evidently these are considerations the nature of which may change greatly during the work of a single generation. We may instance the development by PEARSON of a very extensive system of skew curves, the elaboration of a method of calculating their parameters, and the preparation of the necessary tables, a body of work which has enormously extended the power of modern statistical practice, and which has been, by pertinacity and inspiration alike, practically the work of a single man. Nor is the introduction of the Pearsonian system of frequency curves the only contribution which their author has made to the solution of problems of specification: of even greater importance is the introduction of an objective criterion of goodness of fit. For empirical as the specification of the hypothetical population may be, this empiricism is cleared of its dangers if we can apply a rigorous and objective test of the adequacy with which the proposed population represents the whole of the available facts. Once a statistic, suitable for applying such a test, has been chosen, the exact form of its distribution in random samples must be investigated, in order that we may evaluate the probability that a worse fit should be obtained from a random sample of a population of the type considered. The possibility of developing complete and self-contained tests of goodness of fit deserves very careful consideration, since therein lies our justification for the free use which is made of empirical frequency formulæ. Problems of distribution of great mathematical difficulty have to be faced in this direction.

Although problems of estimation and of distribution may be studied separately, they are intimately related in the development of statistical methods. Logically problems of distribution should have prior consideration, for the study of the random distribution of different suggested statistics, derived from samples of a given size, must guide us in the choice of which statistic it is most profitable to calculate. The fact is, however, that very little progress has been made in the study of the distribution of statistics derived from samples. In 1900 PEARSON (15) gave the exact form of the distribution of $\chi^2$, the Pearsonian test of goodness of fit, and in 1915 the same author published (18) a similar result of more general scope, valid when the observations are regarded as subject to linear constraints. By an easy adaptation (17) the tables of probability derived from this formula may be made available for the more numerous cases in which linear con-

straints are imposed upon the hypothetical population by the means which we employ in its reconstruction. The distribution of the mean of samples of $n$ from a normal population has long been known, but in 1908 "Student" (4) broke new ground by calculating the distribution of the ratio which the deviation of the mean from its population value bears to the standard deviation calculated from the sample. At the same time he gave the exact form of the distribution in samples of the standard deviation. In 1915 FISHER (5) published the curve of distribution of the correlation coefficient for the standard method of calculation, and in 1921 (6) he published the corresponding series of curves for intraclass correlations. The brevity of this list is emphasised by the absence of investigation of other important statistics, such as the regression coefficients, multiple correlations, and the correlation ratio. A formula for the probable error of any statistic is, of course, a practical necessity, if that statistic is to be of service: and in the majority of cases such formulæ have been found, chiefly by the labours of PEARSON and his school, by a first approximation, which describes the distribution with sufficient accuracy if the sample is sufficiently large. Problems of distribution, other than the distribution of statistics, used to be not uncommon as examination problems in probability, and the physical importance of problems of this type may be exemplified by the chemical laws of mass action, by the statistical mechanics of GIBBS, developed by JEANS in its application to the theory of gases, by the electron theory of LORENTZ, and by PLANCK'S development of the theory of quanta, although in all these applications the methods employed have been, from the statistical point of view, relatively simple.

The discussions of theoretical statistics may be regarded as alternating between problems of estimation and problems of distribution. In the first place a method of calculating one of the population parameters is devised from common-sense considerations: we next require to know its probable error, and therefore an approximate solution of the distribution, in samples, of the statistic calculated. It may then become apparent that other statistics may be used as estimates of the same parameter. When the probable errors of these statistics are compared, it is usually found that, in large samples, one particular method of calculation gives a result less subject to random errors than those given by other methods of calculation. Attacking the problem more thoroughly, and calculating the surface of distribution of any two statistics, we may find that the whole of the relevant information contained in one is contained in the other: or, in other words, that when once we know the other, knowledge of the first gives us no further information as to the value of the parameter. Finally it may be possible to prove, as in the case of the Mean Square Error, derived from a sample of normal population (7), that a particular statistic summarises the whole of the information relevant to the corresponding parameter, which the sample contains. In such a case the problem of estimation is completely solved.

## 4. Criteria of Estimation.

The common-sense criterion employed in problems of estimation may be stated thus :— That when applied to the whole population the derived statistic should be equal to the parameter. This may be called the *Criterion of Consistency*. It is often the only test applied : thus, in estimating the standard deviation of a normally distributed population, from an ungrouped sample, either of the two statistics—

$$\sigma_1 = \frac{1}{n}\sqrt{\frac{\pi}{2}}\,\mathrm{S}\,(|x-\bar{x}|) \qquad \text{(Mean error)}$$

and

$$\sigma_2 = \sqrt{\frac{1}{n}\mathrm{S}\,(x-\bar{x})^2} \qquad \text{(Mean square error)}$$

will lead to the correct value, $\sigma$, when calculated from the whole population. They both thus satisfy the criterion of consistency, and this has led many computers to use the first formula, although the result of the second has 14 per cent. greater weight (7), and the labour of increasing the number of observations by 14 per cent. can seldom be less than that of applying the more accurate formula.

Consideration of the above example will suggest a second criterion, namely :—That in large samples, when the distributions of the statistics tend to normality, that statistic is to be chosen which has the least probable error.

This may be called the *Criterion of Efficiency*. It is evident that if for large samples one statistic has a probable error double that of a second, while both are proportional to $n^{-\frac{1}{2}}$, then the first method applied to a sample of $4n$ values will be no more accurate than the second applied to a sample of any $n$ values. If the second method makes use of the whole of the information available, the first makes use of only one-quarter of it, and its efficiency may therefore be said to be 25 per cent. To calculate the efficiency of any given method, we must therefore know the probable error of the statistic calculated by that method, and that of the most efficient statistic which could be used. The square of the ratio of these two quantities then measures the efficiency.

The criterion of efficiency is still to some extent incomplete, for different methods of calculation may tend to agreement for large samples, and yet differ for all finite samples. The complete criterion suggested by our work on the mean square error (7) is :—

That the statistic chosen should summarise the whole of the relevant information supplied by the sample.

This may be called the *Criterion of Sufficiency*.

In mathematical language we may interpret this statement by saying that if $\theta$ be the parameter to be estimated, $\theta_1$ a statistic which contains the whole of the information as to the value of $\theta$, which the sample supplies, and $\theta_2$ any other statistic, then the

surface of distribution of pairs of values of $\theta_1$ and $\theta_2$, for a given value of $\theta$, is such that for a given value of $\theta_1$, the distribution of $\theta_2$ does not involve $\theta$. In other words, when $\theta_1$ is known, knowledge of the value of $\theta_2$ throws no further light upon the value of $\theta$.

It may be shown that a statistic which fulfils the criterion of sufficiency will also fulfil the criterion of efficiency, when the latter is applicable. For, if this be so, the distribution of the statistics will in large samples be normal, the standard deviations being proportional to $n^{-\frac{1}{2}}$. Let this distribution be

$$df = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\, e^{-\frac{1}{1-r^2}\left\{\frac{\overline{\theta_1-\theta}^2}{2\sigma_1^2} - \frac{2r\overline{\theta_1-\theta}\,\overline{\theta_2-\theta}}{2\sigma_1\sigma_2} + \frac{\overline{\theta_2-\theta}^2}{2\sigma_2^2}\right\}} d\theta_1\, d\theta_2,$$

then the distribution of $\theta_1$ is

$$df = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{\overline{\theta_1-\theta}^2}{2\sigma_1^2}}\, d\theta_1,$$

so that for a given value of $\theta_1$ the distribution of $\theta_2$ is

$$df = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\left\{\frac{r\overline{\theta_1-\theta}}{\sigma_1} - \frac{\overline{\theta_2-\theta}}{\sigma_2}\right\}^2} d\theta_2;$$

and if this does not involve $\theta$, we must have

$$r\sigma_2 = \sigma_1;$$

showing that $\sigma_1$ is necessarily less than $\sigma_2$, and that the efficiency of $\theta_2$ is measured by $r^2$, when $r$ is its correlation in large samples with $\theta_1$.

Besides this case we shall see that the criterion of sufficiency is also applicable to finite samples, and to those cases when the weight of a statistic is not proportional to the number of the sample from which it is calculated.

## 5. Examples of the Use of the Criterion of Consistency.

In certain cases the criterion of consistency is sufficient for the solution of problems of estimation. An example of this occurs when a fourfold table is interpreted as representing the double dichotomy of a normal surface. In this case the dichotomic ratios of the two variates, together with the correlation, completely specify the four fractions into which the population is divided. If these are equated to the four fractions into which the sample is divided, the correlation is determined uniquely.

In other cases where a small correction has to be made, the amount of the correction is not of sufficient importance to justify any great refinement in estimation, and it is sufficient to calculate the discrepancy which appears when the uncorrected method is applied to the whole population. Of this nature is Sheppard's correction for grouping,

and it will illustrate this use of the criterion of consistency if we derive formulæ for this correction without approximation.

Let $\xi$ be the value of the variate at the mid point of any group, $a$ the interval of grouping, and $x$ the true value of the variate at any point, then the $k^{\text{th}}$ moment of an infinite grouped sample is

$$\sum_{p=-\infty}^{p=\infty} \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \xi^k f(x)\, dx,$$

in which of $f(x)\,dx$ is the frequency, in any element $dx$, of the ungrouped population, and

$$\xi = \left(p + \frac{\theta}{2\pi}\right) a,$$

$p$ being any integer.

Evidently the $k^{\text{th}}$ moment is periodic in $\theta$, we will therefore equate it to

$$A_0 + A_1 \sin\theta + A_2 \sin 2\theta \ldots$$

$$+ B_1 \cos\theta + B_2 \cos 2\theta \ldots.$$

Then

$$A_0 = \frac{1}{2\pi} \sum_{p=-\infty}^{p=\infty} \int_0^{2\pi} d\theta \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \xi^k f(x)\, dx$$

$$A_S = \frac{1}{\pi} \sum_{p=-\infty}^{p=\infty} \int_0^{2\pi} \sin s\theta\, d\theta \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \xi^k f(x)\, dx,$$

$$B_S = \frac{1}{\pi} \sum_{p=-\infty}^{p=\infty} \int_0^{2\pi} \cos s\theta\, d\theta \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \xi^k f(x)\, dx.$$

But

$$\theta = \frac{2\pi}{a}\,\xi - 2\pi p,$$

therefore

$$d\theta = \frac{2\pi}{a}\, d\xi,$$

$$\sin s\theta = \sin \frac{2\pi}{a}\, s\xi,$$

$$\cos s\theta = \cos \frac{2\pi}{a}\, s\xi,$$

hence

$$A_0 = \frac{1}{a} \int_{-\infty}^{\infty} d\xi \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \xi^k f(x)\, dx = \frac{1}{a} \int_{-\infty}^{\infty} f(x)\, dx \int_{x-\frac{1}{2}a}^{x+\frac{1}{2}a} \xi^k\, d\xi$$

Inserting the values 1, 2, 3 and 4 for $k$, we obtain for the aperiodic terms of the four moments of the grouped population

$$
\begin{aligned}
{}_1\mathrm{A}_0 &= \int_{-\infty}^{\infty} x f(x)\, dx, \\
{}_2\mathrm{A}_0 &= \int_{-\infty}^{\infty} \left(x^2 + \frac{a^2}{12}\right) f(x)\, dx, \\
{}_3\mathrm{A}_0 &= \int_{-\infty}^{\infty} \left(x^3 + \frac{a^2 x}{4}\right) f(x)\, dx, \\
{}_4\mathrm{A}_0 &= \int_{-\infty}^{\infty} \left(x^4 + \frac{a^2 x^2}{2} + \frac{a^4}{80}\right) f(x)\, dx.
\end{aligned}
$$

If we ignore the periodic terms, these equations lead to the ordinary SHEPPARD corrections for the second and fourth moment. The nature of the approximation involved is brought out by the periodic terms. In the absence of high contact at the ends of the curve, the contribution of these will, of course, include the terms given in a recent paper by PEARSON (8); but even with high contact it is of interest to see for what degree of coarseness of grouping the periodic terms become sensible.

Now

$$
\begin{aligned}
\mathrm{A}_s &= \frac{1}{\pi} \sum_{p=-\infty}^{p=\infty} \int_0^{2\pi} \sin s\theta\, d\theta \int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} \bar{\xi}^k f(x)\, dx, \\
&= \frac{2}{a} \int_{-\infty}^{\infty} \sin \frac{2\pi s \bar{\xi}}{a}\, d\bar{\xi} \int_{\bar{\xi}-\frac{1}{2}a}^{\bar{\xi}+\frac{1}{2}a} \bar{\xi}^k f(x)\, dx, \\
&= \frac{2}{a} \int_{-\infty}^{\infty} f(x)\, dx \int_{x-\frac{1}{2}a}^{x+\frac{1}{2}a} \bar{\xi}^k \sin \frac{2\pi s \bar{\xi}}{a}\, d\bar{\xi}.
\end{aligned}
$$

But

$$
\frac{2}{a} \int_{x-\frac{1}{2}a}^{x+\frac{1}{2}a} \bar{\xi} \sin \frac{2\pi s \bar{\xi}}{a}\, d\bar{\xi} = -\frac{a}{\pi s} \cos \frac{2\pi s x}{a} \cos \pi s,
$$

therefore

$${}_1\mathrm{A}_s = (-)^{s+1} \frac{a}{\pi s} \int_{-\infty}^{\infty} \cos \frac{2\pi s x}{a}\, f(x)\, dx;$$

similarly the other terms of the different moments may be calculated.

For a normal curve referred to the true mean

$${}_1\mathrm{A}_s = (-)^{s+1} \frac{2\epsilon}{s} e^{-\frac{s^2\sigma^2}{2\epsilon^2}},$$

$${}_1\mathrm{B}_s = 0,$$

in which

$$a = 2\pi\epsilon.$$

The error of the mean is therefore

$$-2\epsilon\left(e^{-\frac{\sigma^2}{2\epsilon^2}} \sin\theta - \tfrac{1}{2} e^{-\frac{4\sigma^2}{2\epsilon^2}} \sin 2\theta + \tfrac{1}{3} e^{-\frac{9\sigma^2}{2\epsilon^2}} \sin 3\theta - \ldots\right).$$

To illustrate a coarse grouping, take the group interval equal to the standard deviation : then

$$\epsilon = \frac{\sigma}{2\pi},$$

and the error is

$$-\frac{\sigma}{\pi} e^{-2\pi^2} \sin\theta$$

with sufficient accuracy. The standard error of the mean being $\frac{\sigma}{\sqrt{n}}$, we may calculate the size of the sample for which the error due to the periodic terms becomes equal to one-tenth of the standard error, by putting

$$\frac{\sigma}{10\sqrt{n}} = \frac{\sigma}{\pi} e^{-2\pi^2},$$

whence

$$n = \frac{\pi^2}{100} e^{4\pi^2} = 13{,}790 \text{ billion.}$$

For the second moment

$$\mathrm{B}_s = (-)^s 4\left(\sigma^2 + \frac{\epsilon^2}{s^2}\right) e^{-\frac{s^2\sigma^2}{2\epsilon^2}},$$

and, if we put

$$\frac{\sqrt{2}\sigma^2}{10\sqrt{n}} = 4\sigma^2 e^{-2\pi^2},$$

there results

$$n = \tfrac{1}{800} e^{4\pi^2} = 175 \text{ billion.}$$

The error, while still very minute, is thus more important for the second than for the first moment.

For the third moment

$$\mathrm{A}_s = (-)^s \frac{6\sigma^4 s}{\epsilon}\left\{1 + \frac{\epsilon^2}{s^2\sigma^2} - \frac{\epsilon^4}{3s^4\sigma^4}(\pi^2 s^2 - 6)\right\} e^{-\frac{s^2\sigma^2}{2\epsilon^2}};$$

putting

$$\frac{\sqrt{15}\sigma^3}{10\sqrt{n}} = 12\pi\sigma^3 e^{-2\pi},$$

$$n = \frac{1}{960\pi^2} e^{4\pi^2} = 14{\cdot}7 \text{ billion.}$$

While for the fourth moment

$$\mathrm{B}_s = (-)^{s+1} \frac{8\sigma^6 s^2}{\epsilon^2}\left\{1 - (\pi^2 s^2 - 3)\frac{\epsilon^4}{s^4\sigma^4} - (\pi^2 s^2 - 6)\frac{\epsilon^6}{s^6\sigma^6}\right\} e^{-\frac{s^2\sigma^2}{2\epsilon^2}},$$

so that, if we put,

$$\frac{\sqrt{96}\sigma^4}{10\sqrt{n}} = 32\pi^2\sigma^4 e^{-2\pi^2},$$

$$n = \frac{3}{3200\pi^4} e^{4\pi^2} = 1{\cdot}34 \text{ billion.}$$

In a similar manner the exact form of SHEPPARD's correction may be found for other curves ; for the normal curve we may say that the periodic terms are exceedingly minute so long as $a$ is less than $\sigma$, though they increase very rapidly if $a$ is increased beyond this point. They are of increasing importance as higher moments are used, not only absolutely, but relatively to the increasing probable errors of the higher moments. The principle upon which the correction is based is merely to find the error when the moments are calculated from an infinite grouped sample ; the corrected moment therefore fulfils the criterion of consistency, and so long as the correction is small no greater refinement is required.

Perhaps the most extended use of the criterion of consistency has been developed by PEARSON in the "Method of Moments." In this method, which is without question of great practical utility, different forms of frequency curves are fitted by calculating as many moments of the sample as there are parameters to be evaluated. The parameters chosen are those of an infinite population of the specified type having the same moments as those calculated from the sample.

The system of curves developed by PEARSON has four variable parameters, and may be fitted by means of the first four moments. For this purpose it is necessary to confine attention to curves of which the first four moments are finite ; further, if the accuracy of the fourth moment should increase with the size of the sample, that is, if its probable error should not be infinitely great, the first eight moments must be finite. This restriction requires that the class of distribution in which this condition is not fulfilled should be set aside as "heterotypic," and that the fourth moment should become practically valueless as this class is approached. It should be made clear, however, that there is nothing anomalous about these so-called "heterotypic" distributions except the fact that the method of moments cannot be applied to them. Moreover, for that class of distribution to which the method can be applied, it has not been shown, except in the case of the normal curve, that the best values will be obtained by the method of moments. The method will, in these cases, certainly be serviceable in yielding an approximation, but to discover whether this approximation is a good or a bad one, and to improve it, if necessary, a more adequate criterion is required.

A single example will be sufficient to illustrate the practical difficulty alluded to above. If a point P lie at known (unit) distance from a straight line AB, and lines be drawn at random through P, then the distribution of the points of intersection with AB will be distributed so that the frequency in any range $dx$ is

$$df = \frac{1}{\pi} \cdot \frac{dx}{1+(x-m)^2},$$

in which $x$ is the distance of the infinitesimal range $dx$ from a fixed point O on the line, and $m$ is the distance, from this point, of the foot of the perpendicular PM. The distri-

bution will be a symmetrical one (Type VII.) having its centre at $x = m$ (fig. 1). It is therefore a perfectly definite problem to estimate the value of $m$ (to find the best value of $m$) from a random sample of values of $x$. We have stated the problem in its simplest possible form: only one parameter is required, the middle point of the distribution.

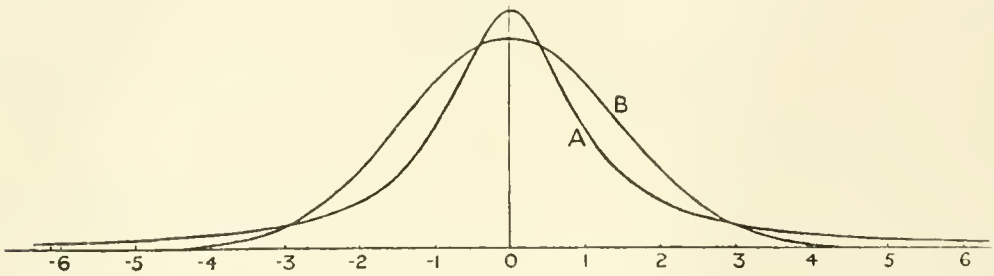



Fig. 1. Symmetrical error curves of equal intrinsic accuracy.

$$\text{A} \quad \ldots\ldots\ldots \quad df = \frac{1}{\pi}\,\frac{dx}{1+x^2}.$$

$$\text{B} \quad \ldots\ldots\ldots \quad df = \frac{1}{2\sqrt{\pi}}\,e^{-\frac{x^2}{4}}.$$

By the method of moments, this should be given by the first moment, that is by the mean of the observations: such would seem to be at least a good estimate. It is, however, entirely valueless. The distribution of the mean of such samples is in fact the same, identically, as that of a single observation. In taking the mean of 100 values of $x$, we are no nearer obtaining the value of $m$ than if we had chosen any value of $x$ out of the 100. The problem, however, is not in the least an impracticable one: clearly from a large sample we ought to be able to estimate the centre of the distribution with some precision; the mean, however, is an entirely useless statistic for the purpose. By taking the median of a large sample, a fair approximation is obtained, for the standard error of the median of a large sample of $n$ is $\frac{\pi}{2\sqrt{n}}$, which, alone, is enough to show that by adopting adequate statistical methods it must be possible to estimate the value for $m$, with increasing accuracy, as the size of the sample is increased.

This example serves also to illustrate the practical difficulty which observers often find, that a few extreme observations appear to dominate the value of the mean. In these cases the rejection of extreme values is often advocated, and it may often happen that gross errors are thus rejected. As a statistical measure, however, the rejection of observations is too crude to be defended: and unless there are other reasons for rejection than mere divergence from the majority, it would be more philosophical to accept these extreme values, not as gross errors, but as indications that the distribution of errors is not normal. As we shall show, the only Pearsonian curve for which the mean

is the best statistic for locating the curve, is the normal or gaussian curve of errors. If the curve is not of this form the mean is not necessarily, as we have seen, of any value whatever. The determination of the true curves of variation for different types of work is therefore of great practical importance, and this can only be done by different workers recording their data in full without rejections, however they may please to treat the data so recorded. Assuredly an observer need be exposed to no criticism, if after recording data which are not probably normal in distribution, he prefers to adopt some value other than the arithmetic mean.

## 6. Formal Solution of Problems of Estimation.

The form in which the criterion of sufficiency has been presented is not of direct assistance in the solution of problems of estimation. For it is necessary first to know the statistic concerned and its surface of distribution, with an infinite number of other statistics, before its sufficiency can be tested. For the solution of problems of estimation we require a method which for each particular problem will lead us automatically to the statistic by which the criterion of sufficiency is satisfied. Such a method is, I believe, provided by the Method of Maximum Likelihood, although I am not satisfied as to the mathematical rigour of any proof which I can put forward to that effect. Readers of the ensuing pages are invited to form their own opinion as to the possibility of the method of the maximum likelihood leading in any case to an insufficient statistic. For my own part I should gladly have withheld publication until a rigorously complete proof could have been formulated; but the number and variety of the new results which the method discloses press for publication, and at the same time I am not insensible of the advantage which accrues to Applied Mathematics from the co-operation of the Pure Mathematician, and this co-operation is not infrequently called forth by the very imperfections of writers on Applied Mathematics.

If in any distribution involving unknown parameters $\theta_1, \theta_2, \theta_3, \ldots$, the chance of an observation falling in the range $dx$ be represented by

$$f(x, \theta_1, \theta_2, \ldots)\, dx,$$

then the chance that in a sample of $n$, $n_1$ fall in the range $dx_1$, $n_2$ in the range $dx_2$, and so on, will be

$$\frac{n!}{\Pi(n_p!)} \Pi \{f(x_p, \theta_1, \theta_2, \ldots)\, dx_p\}^{n_p}.$$

The method of maximum likelihood consists simply in choosing that set of values for the parameters which makes this quantity a maximum, and since in this expression the parameters are only involved in the function $f$, we have to make

$$\mathrm{S}(\log f)$$

a maximum for variations of $\theta_1$, $\theta_2$, $\theta_3$, &c. In this form the method is applicable to the fitting of populations involving any number of variates, and equally to discontinuous as to continuous distributions.

In order to make clear the distinction between this method and that of BAYES, we will apply it to the same type of problem as that which BAYES discussed, in the hope of making clear exactly of what kind is the information which a sample is capable of supplying. This question naturally first arose, not with respect to populations distributed in frequency curves and surfaces, but with respect to a population regarded as divided into two classes only, in fact in problems of *probability*. A certain proportion, $p$, of an infinite population is supposed to be of a certain kind, *e.g.*, "successes," the remainder are then "failures." A sample of $n$ is taken and found to contain $x$ successes and $y$ failures. The chance of obtaining such a sample is evidently

$$\frac{n!}{x!\,y!}\,p^x(1-p)^y.$$

Applying the method of maximum likelihood, we have

$$\mathrm{S}(\log f) = x\log\hat{p} + y\log(1-\hat{p})$$

whence, differentiating with respect to $p$, in order to make this quantity a maximum,

$$\frac{x}{p} = \frac{y}{1-\hat{p}}, \quad \text{or} \quad \hat{p} = \frac{x}{n}.$$

The question then arises as to the accuracy of this determination. This question was first discussed by BAYES (10), in a form which we may state thus. After observing this sample, when we know $p$, what is the *probability* that $p$ lies in any range $dp$? In other words, what is the frequency distribution of the values of $p$ in populations which are selected by the restriction that a sample of $n$ taken from each of them yields $x$ successes. Without further data, as BAYES perceived, this problem is insoluble. To render it capable of mathematical treatment, BAYES introduced the *datum*, that among the populations upon which the experiment was tried, those in which $p$ lay in the range $dp$ were equally frequent for all equal ranges $dp$. The probability that the value of $p$ lay in any range $dp$ was therefore assumed to be simply $dp$, before the sample was taken. After the selection effected by observing the sample, the probability is clearly proportional to

$$p^x(1-p)^y\,dp.$$

After giving this solution, based upon the particular datum stated, BAYES adds a *scholium* the purport of which would seem to be that in the absence of all knowledge save that supplied by the sample, it is reasonable to assume this particular *a priori* distribution of $p$. The *result*, the *datum*, and the *postulate* implied by the *scholium*, have all been somewhat loosely spoken of as BAYES' Theorem.

The postulate would, if true, be of great importance in bringing an immense variety of questions within the domain of probability. It is, however, evidently extremely arbitrary. Apart from evolving a vitally important piece of knowledge, that of the exact form of the distribution of values of $p$, out of an assumption of complete ignorance, it is not even a unique solution. For we might never have happened to direct our attention to the particular quantity $p$: we might equally have measured probability upon an entirely different scale. If, for instance,

$$\sin\theta = 2p-1,$$

the quantity, $\theta$, measures the degree of probability, just as well as $p$, and is even, for some purposes, the more suitable variable. The chance of obtaining a sample of $x$ successes and $y$ failures is now

$$\frac{n!}{2^n x!\, y!}(1+\sin\theta)^x (1-\sin\theta)^y;$$

applying the method of maximum likelihood,

$$\mathrm{S}(\log f) = x\log(1+\sin\theta) + y\log(1-\sin\theta) - n\log 2,$$

and differentiating with respect to $\theta$,

$$\frac{x\cos\theta}{1+\sin\theta} = \frac{y\cos\theta}{1-\sin\theta}, \qquad \text{whence} \qquad \sin\theta = \frac{x-y}{2n},$$

an exactly equivalent solution to that obtained using the variable $p$. But what *a priori* assumption are we to make as to the distribution of $\theta$? Are we to assume that $\theta$ is equally likely to lie in all equal ranges $d\theta$? In this case the *a priori* probability will be $d\theta/\pi$, and that after making the observations will be proportional to

$$(1+\sin\theta)^x (1-\sin\theta)^y\, d\theta.$$

But if we interpret this in terms of $p$, we obtain

$$p^x(1-p)^y \frac{dp}{\sqrt{p(1-p)}} = p^{x-\frac{1}{2}}(1-p)^{y-\frac{1}{2}}\, dp,$$

a result inconsistent with that obtained previously. In fact, the distribution previously assumed for $p$ was equivalent to assuming the special distribution for $\theta$,

$$df = \frac{\cos\theta}{2}\, d\theta,$$

the arbitrariness of which is fully apparent when we use any variable other than $p$.

In a less obtrusive form the same species of arbitrary assumption underlies the method

known as that of inverse probability. Thus, if the same observed result A might be the consequence of one or other of two hypothetical conditions X and Y, it is assumed that the probabilities of X and Y are in the same ratio as the probabilities of A occurring on the two assumptions, X is true, Y is true. This amounts to assuming that before A was observed, it was known that our universe had been selected at random for an infinite population in which X was true in one half, and Y true in the other half. Clearly such an assumption is entirely arbitrary, nor has any method been put forward by which such assumptions can be made even with consistent uniqueness. There is nothing to prevent an irrelevant distinction being drawn among the hypothetical conditions represented by X, so that we have to consider two hypothetical possibilities $X_1$ and $X_2$, on both of which A will occur with equal frequency. Such a distinction should make no difference whatever to our conclusions ; but on the principle of inverse probability it does so, for if previously the relative probabilities were reckoned to be in the ratio $x$ to $y$, they must now be reckoned $2x$ to $y$. Nor has any criterion been suggested by which it is possible to separate such irrelevant distinctions from those which are relevant.

There would be no need to emphasise the baseless character of the assumptions made under the titles of inverse probability and Bayes' Theorem in view of the decisive criticism to which they have been exposed at the hands of Boole, Venn, and Chrystal, were it not for the fact that the older writers, such as Laplace and Poisson, who accepted these assumptions, also laid the foundations of the modern theory of statistics, and have introduced into their discussions of this subject ideas of a similar character. I must indeed plead guilty in my original statement of the Method of the Maximum Likelihood (9) to having based my argument upon the principle of inverse probability ; in the same paper, it is true, I emphasised the fact that such inverse probabilities were relative only. That is to say, that while we might speak of one value of $p$ as having an inverse probability three times that of another value of $p$, we might on no account introduce the differential element $dp$, so as to be able to say that it was three times as probable that $p$ should lie in one rather than the other of two equal elements. Upon consideration, therefore, I perceive that the word probability is wrongly used in such a connection : probability is a ratio of frequencies, and about the frequencies of such values we can know nothing whatever. We must return to the actual fact that one value of $p$, of the frequency of which we know nothing, would yield the observed result three times as frequently as would another value of $p$. If we need a word to characterise this relative property of different values of $p$, I suggest that we may speak without confusion of the *likelihood* of one value of $p$ being thrice the likelihood of another, bearing always in mind that likelihood is not here used loosely as a synonym of probability, but simply to express the relative frequencies with which such values of the hypothetical quantity $p$ would in fact yield the observed sample.

The solution of the problems of calculating from a sample the parameters of the hypothetical population, which we have put forward in the method of maximum likeli-

hood, consists, then, simply of choosing such values of these parameters as have the maximum likelihood. Formally, therefore, it resembles the calculation of the mode of an inverse frequency distribution. This resemblance is quite superficial: if the scale of measurement of the hypothetical quantity be altered, the mode must change its position, and can be brought to have any value, by an appropriate change of scale; but the optimum, as the position of maximum likelihood may be called, is entirely unchanged by any such transformation. Likelihood also differs from probability* in that it is not a differential element, and is incapable of being integrated: it is assigned to a particular point of the range of variation, not to a particular element of it. There is therefore an absolute measure of probability in that the unit is chosen so as to make all the elementary probabilities add up to unity. There is no such absolute measure of likelihood. It may be convenient to assign the value unity to the maximum value, and to measure other likelihoods by comparison, but there will then be an infinite number of values whose likelihood is greater than one-half. The sum of the likelihoods of admissible values will always be infinite.

Our interpretation of BAYES' problem, then, is that the likelihood of any value of $p$ is proportional to

$$p^x (1-p)^y,$$

and is therefore a maximum when

$$p = \frac{x}{n},$$

which is the best value obtainable from the sample; we shall term this the *optimum* value of $p$. Other values of $p$ for which the likelihood is not much less cannot, however, be deemed unlikely values for the true value of $p$. We do not, and cannot, know, from the information supplied by a sample, anything about the probability that $p$ should lie between any named values.

The reliance to be placed on such a result must depend upon the frequency distribution of $x$, in different samples from the same population. This is a perfectly objective statistical problem, of the kind we have called problems of distribution; it is, however, capable of an approximate solution, directly from the mathematical form of the likelihood.

When for large samples the distribution of any statistic, $\theta_1$, tends to normality, we

* It should be remarked that likelihood, as above defined, is not only fundamentally distinct from mathematical probability, but also from the logical "probability" by which Mr. KEYNES (21) has recently attempted to develop a method of treatment of uncertain inference, applicable to those cases where we lack the statistical information necessary for the application of mathematical probability. Although, in an important class of cases, the likelihood may be held to measure the degree of our rational belief in a conclusion, in the same sense as Mr. KEYNES' "probability," yet since the latter quantity is constrained, somewhat arbitrarily, to obey the addition theorem of mathematical probability, the likelihood is a quantity which falls definitely outside its scope.

may write down the chance for a given value of the parameter $\theta$, that $\theta_1$ should lie in the range $d\theta_1$ in the form

$$\Phi = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\theta_1-\theta)^2}{2\sigma^2}} d\theta_1.$$

The mean value of $\theta_1$ will be the true value $\theta$, and the standard deviation is $\sigma$, the sample being assumed sufficiently large for us to disregard the dependence of $\sigma$ upon $\theta$.

The likelihood of any value, $\theta$, is proportional to

$$e^{-\frac{(\theta_1-\theta)^2}{2\sigma^2}},$$

this quantity having its maximum value, unity, when

$$\theta = \theta_1;$$

for

$$\frac{\partial}{\partial\theta}\log\Phi = \frac{\theta_1-\theta}{\sigma^2}.$$

Differentiating now a second time

$$\frac{\partial^2}{\partial\theta^2}\log\Phi = -\frac{1}{\sigma^2}.$$

Now $\Phi$ stands for the total frequency of all samples for which the chosen statistic has the value $\theta_1$, consequently $\Phi = \mathrm{S}'(\phi)$, the summation being taken over all such examples, where $\phi$ stands for the probability of occurrence of a certain specified sample. For which we know that

$$\log\phi = \mathrm{C} + \mathrm{S}(\log f),$$

the summation being taken over the individual members of the sample.

If now we expand $\log f$ in the form

$$\log f(\theta) = \log f(\theta_1) + \overline{\theta-\theta_1}\frac{\partial}{\partial\theta}\log f(\theta_1) + \frac{\overline{\theta-\theta_1}^2}{\underline{2}}\frac{\partial^2}{\partial\theta^2}\log f(\theta_1) + \dots,$$

or

$$\log f = \log f_1 + a\,\overline{\theta-\theta_1} + \frac{b}{2}\overline{\theta-\theta_1}^2 + \dots,$$

we have

$$\log\phi = \mathrm{C} + \overline{\theta-\theta_1}\,\mathrm{S}(a) + \tfrac{1}{2}\,\overline{\theta-\theta_1}^2\,\mathrm{S}(b) + \dots;$$

now for optimum statistics

$$\mathrm{S}(a) = 0,$$

and for sufficiently large samples $\mathrm{S}(b)$ differs from $nb$ only by a quantity of order $\sqrt{n}\,\sigma_b$; moreover, $\theta-\theta_1$ being of order $n^{-\frac{1}{2}}$, the only terms in $\log\phi$ which are not reduced without limit, as $n$ is increased, are

$$\log\phi = \mathrm{C} + \tfrac{1}{2}n\,\bar{b}\,\overline{\theta-\theta_1}^2;$$

hence

$$\phi \propto e^{\frac{1}{2}n\,\bar{b}\,\overline{\theta-\theta_1}^2}.$$

Now this factor is constant for all samples which have the same value of $\theta_1$, hence the variation of $\Phi$ with respect to $\theta$ is represented by the same factor, and consequently

$$\log \Phi = C' + \tfrac{1}{2}n\,\bar{b}\,\overline{\theta-\theta_1}^2;$$

whence

$$-\frac{1}{\sigma_{\theta_1}^2} = \frac{\partial^2}{\partial\theta^2}\log\Phi = n\,\bar{b},$$

where

$$b = \frac{\partial^2}{\partial\theta^2}\log f(\theta_1),$$

$\theta_1$ being the optimum value of $\theta$.

The formula

$$-\frac{1}{\sigma_\theta^2} = x\,\overline{\frac{\partial^2}{\partial\theta^2}\log f}$$

supplies the most direct way known to me of finding the probable errors of statistics. It may be seen that the above proof applies only to statistics obtained by the method of maximum likelihood.*

For example, to find the standard deviation of

$$\hat{p} = \frac{x}{n}$$

* A similar method of obtaining the standard deviations and correlations of statistics derived from large samples was developed by Pearson and Filon in 1898 (16). It is unfortunate that in this memoir no sufficient distinction is drawn between the *population* and the *sample*, in consequence of which the formulæ obtained indicate that the likelihood is always a maximum (for continuous distributions) when the *mean* of each variate in the sample is equated to the corresponding mean in the population (16, p. 232, "$A_r = 0$"). If this were so the mean would always be a sufficient statistic for location; but as we have already seen, and will see later in more detail, this is far from being the case. The same argument, indeed, is applied to all statistics, as to which nothing but their *consistency* can be truly affirmed.

The probable errors obtained in this way are those appropriate to the method of maximum likelihood, but not in other cases to statistics obtained by the method of moments, by which method the examples given were fitted. In the 'Tables for Statisticians and Biometricians' (1914), the probable errors of the constants of the Pearsonian curves are those proper to the method of moments; no mention is there made of this change of practice, nor is the publication of 1898 referred to.

It would appear that shortly before 1898 the process which leads to the correct value, of the probable errors of *optimum* statistics, was hit upon and found to agree with the probable errors of statistics found by the method of moments for *normal* curves and surfaces; without further enquiry it would appear to have been assumed that this process was valid in all cases, its directness and simplicity being peculiarly attractive. The mistake was at that time, perhaps, a natural one; but that it should have been discovered and corrected without revealing the inefficiency of the method of moments is a very remarkable circumstance.

In 1903 the correct formulæ for the probable errors of statistics found by the method of moments are given in 'Biometrika' (19); references are there given to Sheppard (20), whose method is employed, as well as to Pearson and Filon (16), although both the method and the results differ from those of the latter.

in samples from an infinite population of which the true value is $p$,

$$\log f = \log p + y \log (1-p),$$

$$\frac{\partial}{\partial p} \log f = \frac{x}{p} - \frac{y}{1-p},$$

$$\frac{\partial^2}{\partial p^2} \log f = -\frac{x}{p^2} - \frac{y}{(1-p)^2}.$$

Now the mean value of $x$ in $pn$, and of $y$ is $(1-p)n$, hence the mean value of $\frac{\partial^2}{\partial p^2} \log f$ is

$$-\left(\frac{1}{p} + \frac{1}{1-p}\right) n;$$

therefore

$$\sigma_p^2 = \frac{p(1-p)}{n},$$

the well-known formula for the standard error of $p$.

## 7. Satisfaction of the Criterion of Sufficiency.

That the criterion of sufficiency is generally satisfied by the solution obtained by the method of maximum likelihood appears from the following considerations.

If the individual values of any sample of data are regarded as co-ordinates in hyperspace, then any sample may be represented by a single point, and the frequency distribution of an infinite number of random samples is represented by a density distribution in hyperspace. If any set of statistics be chosen to be calculated from the samples, certain regions will provide identical sets of statistics ; these may be called *isostatistical* regions. For any particular space element, corresponding to an actual sample, there will be a particular set of parameters for which the frequency in that element is a maximum ; this will be the optimum set of parameters for that element. If now the set of statistics chosen are those which give the optimum values of the parameters, then all the elements of any part of the same isostatistical region will contain the greatest possible frequency for the same set of values of the parameters, and therefore any region which lies wholly within an isostatistical region will contain its maximum frequency for that set of values.

Now let $\theta$ be the value of any parameter, $\hat{\theta}$ the statistic calculated by the method of maximum likelihood, and $\theta_1$ any other statistic designed to estimate the value of $\theta$, then for a sample of given size, we may take

$$f(\theta, \hat{\theta}, \theta_1)\, d\hat{\theta}\, d\theta_1$$

to represent the frequency with which $\hat{\theta}$ and $\theta_1$ lie in the assigned ranges $d\hat{\theta}$ and $d\theta_1$.

The region $d\hat{\theta}\, d\theta_1$ evidently lies wholly in the isostatistical region $d\hat{\theta}$. Hence the equation

$$\frac{\partial}{\partial \theta} \log f(\theta, \hat{\theta}, \theta_1) = 0$$

is satisfied, irrespective of $\theta_1$, by the value $\theta = \hat{\theta}$. This condition is satisfied if

$$f(\theta, \hat{\theta}, \theta_1) = \phi(\theta, \hat{\theta}) \cdot \phi'(\hat{\theta}, \theta_1);$$

for then

$$\frac{\partial}{\partial \theta} \log f = \frac{\partial}{\partial \theta} \log \phi,$$

and the equation for the optimum degenerates into

$$\frac{\partial}{\partial \theta} \log \phi(\theta, \hat{\theta}) = 0,$$

which does not involve $\theta_1$.

But the factorisation of $f$ into factors involving $(\theta, \hat{\theta})$ and $(\hat{\theta}, \theta_1)$ respectively is merely a mathematical expression of the condition of sufficiency; and it appears that any statistic which fulfils the condition of sufficiency must be a solution obtained by the method of the optimum.

It may be expected, therefore, that we shall be led to a sufficient solution of problems of estimation in general by the following procedure. Write down the formula for the probability of an observation falling in the range $dx$ in the form

$$f(\theta, x)\, dx,$$

where $\theta$ is an unknown parameter. Then if

$$\mathrm{L} = \mathrm{S}(\log f),$$

the summation being extended over the observed sample, L differs by a constant only from the logarithm of the likelihood of any value of $\theta$. The most likely value, $\hat{\theta}$, is found by the equation

$$\frac{\partial \mathrm{L}}{\partial \theta} = 0,$$

and the standard deviation of $\hat{\theta}$, by a second differentiation, from the formula

$$\frac{\partial^2 \mathrm{L}}{\partial \theta^2} = -\frac{1}{\sigma_{\hat{\theta}}^2};$$

this latter formula being applicable only where $\hat{\theta}$ is normally distributed, as is often the case with considerable accuracy in large samples. The value $\sigma_{\hat{\theta}}$ so found is in these cases the least possible value for the standard deviation of a statistic designed to

estimate the same parameter ; it may therefore be applied to calculate the efficiency of any other such statistic.

When several parameters are determined simultaneously, we must equate the second differentials of L, with respect to the parameters, to the coefficients of the quadratic terms in the index of the normal expression which represents the distribution of the corresponding statistics. Thus with two parameters,

$$\frac{\partial^2 \mathrm{L}}{\partial \theta_1^2} = -\frac{1}{1-r^2_{\hat{\theta}_1\hat{\theta}_2}} \cdot \frac{1}{\sigma^2_{\hat{\theta}_1}}, \qquad \frac{\partial^2 \mathrm{L}}{\partial \theta_2^2} = -\frac{1}{1-r^2_{\hat{\theta}_1\hat{\theta}_2}} \cdot \frac{1}{\sigma^2_{\hat{\theta}_2}},$$

$$\frac{\partial^2 \mathrm{L}}{\partial \theta_1 \partial \theta_2} = +\frac{1}{1-r^2_{\hat{\theta}_1\hat{\theta}_2}} \cdot \frac{r}{\sigma_{\hat{\theta}_1}\sigma_{\hat{\theta}_2}},$$

or, in effect, $\sigma_{\hat{\theta}}^2$ is found by dividing the Hessian determinant of L, with respect to the parameters, into the corresponding minor.

The application of these methods to such a series of parameters as occur in the specification of frequency curves may best be made clear by an example.

## 8. The Efficiency of the Method of Moments in Fitting Curves of the Pearsonian Type III.

Curves of Pearson's Type III. offer a good example for the calculation of the efficiency of the Method of Moments. The chance of an observation falling in the range $dx$ is

$$df = \frac{1}{a\,.\,p\,!} \cdot \left(\frac{x-m}{a}\right)^p e^{-\frac{x-m}{a}}\, dx.*$$

By the method of moments the curve is located by means of the statistic $\mu_1$, its dimensions are ascertained from the second moment $\mu_2$, and the remaining parameter $p$ is determined from $\beta_1$. Considering first the problem of location, if $a$ and $p$ were known and we had only to determine $m$, we should take, according to the method of moments,

$$\mu_1 = m_\mu + a\,(p+1),$$

where $m_\mu$ represents the estimate of the parameter $m$, obtained by using the method of moments. The variance of $m_\mu$ is, therefore,

$$\sigma^2_{m_\mu} = \sigma^2_{\mu_1} = \frac{\mu_2}{n} = \frac{a^2\,(p+1)}{n}.$$

If, on the other hand, we aim at greater accuracy, and make the likelihood of the sample a maximum for variations of $m$, we have

$$\mathrm{L} = -n \log a - n \log (p\,!) + p\mathrm{S}\left(\log \frac{x-m}{a}\right) - \mathrm{S}\left(\frac{x-m}{a}\right),$$

* The expression, $x\,!$, is used here and throughout as equivalent to the Gaussian $\Pi\,(x)$, or to $\Gamma\,(x+1)$, whether $x$ is an integer or not.

and the equation to determine $m$ is

$$\frac{\partial \mathrm{L}}{\partial m} = -p\mathrm{S}\left(\frac{1}{x-\hat{m}}\right) + \frac{n}{a} = 0; \quad \ldots\ldots\ldots \quad (1)$$

the accuracy of the value so obtained is found from the second differential,

$$\frac{\partial^2 \mathrm{L}}{\partial m^2} = -p\mathrm{S}\left(\frac{1}{x-m^2}\right),$$

of which the mean value is

$$-\frac{n}{a^2(p-1)},$$

whence

$$\sigma_{\hat{m}}^2 = \frac{a^2(p-1)}{n}.$$

We now see that the efficiency of location by the method of moments is

$$\frac{p-1}{p+1} = 1 - \frac{2}{p+1}.$$

Efficiencies of over 80 per cent. for location are therefore obtained if $p$ exceeds 9; for $p = 1$ the efficiency of location vanishes, as in other cases where the curve makes an angle with the axis at the end of its range.

Turning now to the problem of scaling, we have, by the method of moments,

$$\mu_2 = a^2(p+1),$$

whence, knowing $p$, $a$ is obtained. Since

$$\sigma^2_{\mu_2} = \frac{\beta_2 - 1}{n}\mu_2^2,$$

we must have

$$\sigma^2_{a_\mu} = \frac{\beta_2 - 1}{4n}a^2 = \frac{4+3\beta_1}{8n}a^2 = \frac{p+4}{2(p+1)n}a^2;$$

on the other hand, from the value of L, we find the equation

$$\frac{\partial \mathrm{L}}{\partial a} = -\frac{n}{a}(p+1) + \frac{1}{a^2}\mathrm{S}(x-m) = 0, \quad \ldots\ldots\ldots \quad (2)$$

to be solved for $m$ and $a$ as a simultaneous equation with (1); whence

$$\frac{\partial^2 \mathrm{L}}{\partial m\,\partial a} = -\frac{n}{a^2},$$

and

$$\frac{\partial^2 \mathrm{L}}{\partial a^2} = \frac{n}{a^2}(p+1) - \frac{2}{a^3}\mathrm{S}(x-m),$$

of which the mean value is

$$-\frac{n(p+1)}{a^2}.$$

The variance of $a$, determined from this pair of simultaneous equations, is found by dividing

$$\frac{\partial^2 L}{\partial m^2} = -\frac{n}{a^2(p-1)}$$

by the determinant

$$\begin{vmatrix} -\frac{n}{a^2(p-1)} & -\frac{n}{a^2} \\ -\frac{n}{a^2} & -\frac{n(p+1)}{a^2} \end{vmatrix}$$

which reduces to

$$-\frac{2}{p-1}\cdot\frac{n^2}{a^4},$$

whence

$$\sigma_a^2 = \frac{a^2}{2n},$$

and the efficiency of scaling by the method of moments is

$$\frac{p+1}{p+4} = 1-\frac{3}{p+4}.$$

Efficiency of over 80 per cent. for scaling are, therefore, obtained when $p$ exceeds 11. The efficiency of scaling does not, however, vanish for any possible value of $p$, though it tends to zero, as $p$ approaches its limiting value, $-1$.

Lastly, $p$ is found by the method of moments by putting

$$\frac{4}{p+1} = \beta_1.$$

Now

$$\sigma_{\beta_1}^2 = \frac{\beta_1}{n}(4\beta_4 - 24\beta_2 + 36 + 9\beta_1\beta_2 - 12\beta_3 + 35\beta_1),$$

and for curves of Type III,

$$\beta_2 = 3+\tfrac{3}{2}\beta_1,$$

$$\beta_3 = 2\beta_1\beta_2 + 4\beta_1 = \beta_1(3\beta_1+10),$$

$$\beta_4 = \tfrac{5}{2}(\beta_3+2\beta_2) = \tfrac{5}{2}(3\beta_1^2+13\beta_1+6),$$

hence

$$\sigma_{\beta_1}^2 = \frac{3\beta_1}{n}(5\beta_1+4)(\beta_1+4),$$

$$= \frac{\beta_1^2}{n}\cdot\frac{6(p+2)(p+6)}{p+1},$$

whence it follows, since $n$ is large, that

$$\sigma_{p_\mu}{}^2 = \frac{\overline{p+1}^2}{n} \cdot \frac{6\,(p+2)\,(p+6)}{p+1} = \frac{6}{n}\,(p+1)\,(p+2)\,(p+6).$$

From the value of L,

$$\frac{\partial \mathrm{L}}{\partial p} = -n\,\frac{d}{dp}\log\,(p!) + \mathrm{S}\left(\log\frac{x-m}{a}\right),$$

which equation solved for $m$, $a$ and $p$ as a simultaneous equation with (1) and (2), will yield the set of values for the parameters which has the maximum likelihood. To find the variance of the value of $p$, so obtained, observe that

$$\frac{\partial^2 \mathrm{L}}{\partial m\,\partial p} = -\mathrm{S}\left(\frac{1}{x-m}\right),$$

of which the mean value is

$$-\frac{n}{ap},$$

$$\frac{\partial^2 \mathrm{L}}{\partial a\,\partial p} = -\frac{n}{a^2},$$

and

$$\frac{\partial^2 \mathrm{L}}{\partial p^2} = -n\,\frac{d^2}{dp^2}\log\,(p!).$$

The variance of $p$, derived from this set of simultaneous equations, is therefore found by dividing the minor of $\frac{\partial^2 \mathrm{L}}{\partial p^2}$, namely

$$\frac{2}{p-1} \cdot \frac{n^2}{a^4},$$

by the determinant

$$\frac{n^3}{a^4}\begin{vmatrix} \frac{1}{p-1} & 1 & \frac{1}{p} \\ 1 & p+1 & 1 \\ \frac{1}{p} & 1 & \frac{d^2}{dp^2}\log\,(p!) \end{vmatrix} = \frac{n^3}{a^4} \cdot \frac{1}{p-1}\left\{2\,\frac{d^2}{dp^2}\log\,(p!) - \frac{2}{p} + \frac{1}{p^2}\right\};$$

hence

$$\sigma_p{}^2 = \frac{2}{n\left\{2\,\frac{d^2}{dp^2}\log\,(p!) - \frac{2}{p} + \frac{1}{p^2}\right\}}.$$

When $p$ is large,

$$2\,\frac{d^2}{dp^2}\log\,(p!) - \frac{2}{p} + \frac{1}{p^2} = \tfrac{1}{3}\left(\frac{1}{p^3} - \frac{1}{5p^5} + \frac{1}{7p^7} \cdots\right),$$

so that, approximately,

$$\sigma_{\hat{p}}^2 = \frac{6}{n}(p^3+\tfrac{1}{5}p);$$

for large values of $p$, the efficiency of the method of moments is, therefore, approximately

$$\frac{p^3+\frac{1}{5}p}{\overline{p+1}\,\overline{p+2}\,\overline{p+6}}.$$

Efficiencies of over 80 per cent. occur when $p$ exceeds 38·1 ($\beta_1 = 0{\cdot}102$); evidently the method of moments is effective for determining the form of the curve only when it is relatively close to the normal form. For small values of $p$, the above approximation for the efficiency is not adequate. The true values can easily be obtained from the recently published tables of the Trigamma* function (11). The following values are obtained for the integral values of $p$ from 0 to 5.

| $p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Efficiency . . . | 0 | 0·0274 | 0·0871 | 0·1532 | 0·2159 | 0·2727 |

An interesting point which may be resolved at this stage of the enquiry is to find the variance of $m$, when $a$ and $p$ are not known, derived from the above set of simultaneous equations; that is to say, to calculate the accuracy with which the limiting point of the curve is determined; such determinations are often stated as the result of fitting curves of limited range, but their probable errors are seldom, if ever, evaluated. To obtain the greatest possible accuracy with which such a point can be determined we must divide the minor of $\frac{\partial^2 \mathrm{L}}{\partial m^2}$, namely,

$$\frac{n^2}{a^2}\left\{\overline{p+1}\,\frac{d^2}{dp^2}\log(p!)-1\right\}.$$

by

$$\frac{n^3}{a^4}\cdot\frac{1}{p-1}\left\{2\frac{d^2}{dp^2}\log(p!)-\frac{2}{p}+\frac{1}{p^2}\right\},$$

whence

$$\sigma_{\hat{m}}^2 = \frac{a^2}{n}\frac{\overline{p-1}\left\{\overline{p+1}\,\frac{d^2}{dp^2}\log(p!)-1\right\}}{2\frac{d^2}{dp^2}\log(p!)-\frac{2}{p}+\frac{1}{p^2}}.$$

The position of the limiting point will, when $p$ is at all large, evidently be determined with much less accuracy than is the position, as a whole, of a curve of known form and size. Let $n'$ be a multiplier such that the position of the extremity of a curve calculated

* It is sometimes convenient to write $\mathrm{F}(x)$ for $\frac{d^2}{dx^2}\log(x!)$.

from $nn'$ observations will be determined with the same accuracy as the position, as a whole, of a curve of known form and size, can be determined from a sample of $n$ observations when $n$ is large. Then

$$n' = \frac{\overline{p+1}\,\frac{d^2}{dp^2}\log(p!) - 1}{2\frac{d^2}{dp^2}\log(p!) - \frac{2}{p} + \frac{1}{p^2}};$$

but, when $p$ is large,

$$\overline{p+1}\,\frac{d^2}{dp^2}\log(p!) - 1 = \frac{1}{2p}\left(1 - \frac{2}{3p} + \frac{1}{3p^2}\ldots\right)$$

and

$$2\frac{d^2}{dp^2}\log(p!) - \frac{2}{p} + \frac{1}{p^2} = \frac{1}{3p^3}\left(1 - \frac{1}{5p^2}\ldots\right);$$

therefore

$$n' = \tfrac{3}{2}p^2\left(1 - \frac{2}{3p}p + \frac{8}{15p^2}\ldots\right)$$
$$= \tfrac{3}{2}p^2 - p + \tfrac{4}{5}.$$

For large values of $p$ the probable error of the determination of the end-point may be found approximately by multiplying the probable error of location by

$$(p - \tfrac{1}{3})\sqrt{\tfrac{3}{2}}.$$

As $p$ grows smaller, $n'$ diminishes until it reaches unity, when $p = 1$. For values of $p$ less than 1 it would appear that the end-point had a smaller probable error than the probable error of location, but, as a matter of fact, for these values location is determined by the end-point, and as we see from the vanishing of $\sigma_{\hat{m}}$, whether or not $p$ and $a$ are known, when $p = 1$, the weight of the determination from this point onwards increases more rapidly than $n$, as the sample increases. (See Section 10.)

The above method illustrates how it is possible to calculate the variance of any function of the population parameters as estimated from large samples; by comparing this variance with that of the same function estimated by the method of moments, we may find the efficiency of that method for any proposed function. The above examination, in which the determinations of the locus, the scale, and the form of the curve are treated separately, will serve as a general criterion of the application of the method of moments to curves of Type III. Special combinations of the parameters will, however, be of interest in special cases. It may be noted here that by virtue of equation (2) the function of $m + a(p+1)$ is the same, whether determined by moments or by the method of the optimum:

$$m_\mu + a_\mu(p_\mu + 1) = \hat{m} + \hat{a}(\hat{p} + 1).$$

The efficiency of the method of moments in determining this function is therefore 100 per cent. That this function is the abscissa of the mean does not imply 100 per cent. efficiency of location, for the centre of location of these curves is not the mean (see p. 340).

### 9. Location and Scaling of Frequency Curves in General.

The general problem of the location and scaling of curves may now be treated more generally. This is the problem which presents itself with respect to *error curves* of assumed form, when to find the best value of the quantity measured we must *locate* the curve as accurately as possible, and to find the probable error of the result of this process we must, as accurately as possible, estimate its *scale*.

The *form* of the curve may be specified by a function $\phi$, such that

$$df \propto e^{\phi(\xi)}\, d\xi, \text{ when } \xi = \frac{x-m}{a}.$$

In this expression $\phi$ specifies the form of the curve, which is unaltered by variations of $a$ and $m$.

When a sample of $n$ observations has been taken, the likelihood of any combination of values of $a$ and $m$ is

$$\mathrm{L} = \mathrm{C} - n \log a + \mathrm{S}(\phi),$$

whence

$$\frac{\partial \mathrm{L}}{\partial m} = \mathrm{S}\left(\frac{d\phi}{d\xi} \cdot \frac{d\xi}{dm}\right) = -\frac{1}{a}\mathrm{S}(\phi'),$$

since

$$\frac{\partial \xi}{\partial m} = -\frac{1}{a};$$

also

$$\frac{\partial \mathrm{L}}{\partial a} = -\frac{1}{a}\mathrm{S}(\xi\phi') - \frac{n}{a},$$

since

$$\frac{\partial \xi}{\partial a} = -\frac{\xi}{a}.$$

Differentiating a second time,

$$\frac{\partial^2 \mathrm{L}}{\partial m^2} = \frac{1}{a^2}\mathrm{S}(\phi'');$$

therefore

$$\sigma_{\hat{m}}^{\,2} = -\frac{a^2}{n\phi''}.$$

This expression enables us to compare the accuracy of error curves of different form, when the location is performed in each case by the method which yields the minimum error.

*Example* :—The curve

$$df = \frac{1}{\pi}\frac{d\xi}{1+\xi^2}$$

referred to in Section 5 has an infinite standard deviation, but it is not on that account an error curve of zero accuracy, for

$$\phi = -\log(1+\xi^2), \qquad \phi' = -\frac{2\xi}{1+\xi^2}, \qquad \phi'' = -\frac{2(1-\xi^2)}{(1+\xi^2)^2}.$$

Now

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\xi^2}{(1+\xi^2)^3}\,d\xi = \tfrac{1}{4},$$

hence

$$\overline{\phi''} = -\tfrac{1}{2} \qquad \text{and} \qquad \sigma_m^{\,2} = \frac{2a^2}{n}.$$

The quantity,

$$-\frac{\overline{\phi''}}{a^2} = \frac{1}{2a^2},$$

which is the factor by which $n$ is multiplied in calculating the weight of the estimate made from $n$ measurements, may be called the *intrinsic accuracy* of an *error curve.* In the above example we see that errors distributed so that

$$df = \frac{a}{\pi}\frac{dx}{a^2+x^2}$$

have the same intrinsic accuracy as errors distributed according to the normal curve

$$df = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\,dx,$$

provided

$$\sigma^2 = 2a^2.$$

Fig. 1 illustrates two such curves of equal intrinsic accuracy.

Returning now to the general problem in which

$$\mathrm{L} = \mathrm{C} - n\log a + \mathrm{S}(\phi),$$

we have

$$\frac{\partial^2 \mathrm{L}}{\partial m\,\partial a} = \frac{\mathrm{I}}{a^2}\mathrm{S}(\phi' + \xi\phi'') = \frac{1}{a^2}\mathrm{S}(\xi\phi'')$$

and

$$\frac{\partial^2 \mathrm{L}}{\partial a^2} = \frac{1}{a^2}\mathrm{S}(2\xi\phi' + \xi^2\phi'') + \frac{n}{a^2} = \frac{1}{a^2}\mathrm{S}(\xi^2\phi'' - 1).$$

The latter expression will directly give the accuracy with which $a$ is determined only if

$$\frac{\partial^2 \mathrm{L}}{\partial m\,\partial a} = 0,$$

and we can always arrange that this shall be so by subtracting from $\xi$ the quantity

$$\frac{\overline{\xi\phi''}}{\overline{\phi''}}.$$

Thus in a Type III. curve where, referred to the end of the range,

$$\overline{\xi\phi''} = -1, \qquad \overline{\phi''} = -\frac{1}{p-1},$$

instead of

$$\phi = p \log \xi - \xi$$

we must write

$$\phi = p \log \overline{\xi+p-1} - \overline{\xi+p-1};$$

then

$$\phi' = \frac{p}{\xi+p-1} - 1, \qquad \phi'' = -\frac{p}{\overline{\xi+p-1}^2},$$

hence

$$\frac{\partial^2 \mathrm{L}}{\partial a^2} = \frac{1}{a^2}\mathrm{S}\left(\xi^2\phi'' - 1\right)$$

$$= \frac{1}{a^2}\mathrm{S}\left(-p + \frac{2p\overline{p-1}}{\xi+p-1} - \frac{p\overline{p-1}^2}{\overline{\xi+p-1}^2} - 1\right),$$

of which the mean value is

$$\frac{n}{a^2}\left(-p + 2\overline{p-1} - \overline{p-1} - 1\right) = -\frac{2n}{a^2},$$

hence

$$\sigma_a^2 = \frac{a^2}{2n}.$$

For one particular point of origin, therefore, the variations of the abscissa are uncorrelated with those of $a$; this point may be termed the *centre of location.*

*Example :*—To determine the centre of location of the curve of Type IV.,

$$df \propto e^{-\nu \tan^{-1}\xi}\left(1+\xi^2\right)^{-\frac{r+2}{2}}.$$

Here

$$\phi = -\nu \tan^{-1}\xi - \frac{r+2}{2}\log \overline{1+\xi^2},$$

$$\phi' = -\left(\nu + \overline{r+2}\,\xi\right)\overline{1+\xi^2}^{-1},$$

$$\phi'' = \overline{r+2}\;\overline{1+\xi^2}^{-1} + 2\left(\nu\xi - \overline{r+2}\right)\overline{1+\xi^2}^{-2};$$

from these we find

$$\overline{\phi''} = -\frac{\overline{r+1}\;\overline{r+2}\;\overline{r+4}}{\overline{r+4}^2 + \nu^2},$$

$$\overline{\xi\phi''} = \frac{\overline{r+1}\;\overline{r+2}\;\nu}{\overline{r+4}^2 + \nu^2},$$

so that

$$\frac{\overline{\xi\phi''}}{\overline{\phi''}} = -\frac{\nu}{r+4}.$$

The centre of location, therefore, at the distance from the mode,

$$= -\frac{\nu a}{r+4}.$$

*Example* :—Determine the intrinsic accuracy of an error curve of Type IV. and the efficiency of the method of moments in location and scaling.
Since

$$\overline{\phi''} = -\frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{\overline{r+4}^2+\nu^2}.$$

$$\sigma_{\hat{m}}^2 = \frac{\epsilon^2}{n}\cdot\frac{\overline{r+4}^2+\nu^2}{\overline{r+1}\,\overline{r+2}\,\overline{r+4}};$$

and the intrinsic accuracy of the curve is

$$\frac{1}{\epsilon^2}\frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{\overline{r+4}^2+\nu^2};$$

but

$$\sigma^2_{m_\mu} = \frac{\epsilon^2}{n}\cdot\frac{r^2+\nu^2}{r^2\overline{r-1}},$$

therefore the efficiency of the method of moments in location is

$$\frac{r^2\overline{r-1}\,(\overline{r+4}^2+\nu^2)}{\overline{r+1}\,\overline{r+2}\,\overline{r+4}\,(r^2+\nu^2)}. \qquad (3)$$

When $\nu = 0$, we have for curves of Type VII. an efficiency of location

$$1-\frac{6}{\overline{r+1}\,\overline{r+2}}.$$

The efficiency of location of these curves vanishes at $r = 1$, at which value the standard deviation becomes infinite. Although values down to $-1$ give admissible frequency curves, the conventional limit at which curves are reckoned as heterotypic is at $r = 7$. For this value the efficiency is

$$\frac{49}{132}\cdot\frac{121+\nu^2}{49+\nu^2},$$

which varies from 91·67 per cent. for the symmetrical Type VII. curve, to 37·12 per cent. when $\nu\to\infty$ and the curve to Type V.

Turning to the question of scaling, we find

$$\overline{\xi^2\phi''}-1 = -\frac{\overline{r+1}\,(2\overline{r+4}+\nu^2)}{\overline{r+4}^2+\nu^2},$$

whence

$$\overline{\xi^2\phi''}-1-\frac{\overline{\xi\phi''}^2}{\overline{\phi''}} = -\frac{2\,\overline{r+1}}{r+4},$$

and

$$\sigma_a^2 = \frac{\epsilon^2}{n}\cdot\frac{r+4}{2\,\overline{r+1}};$$

the intrinsic accuracy of scaling is therefore independent of $\nu$. Now for these curves

$$\beta_2 = \frac{3\sqrt{r-1}}{r-2\sqrt{r-3}}\left(r+6-\frac{8r^2}{r^2+\nu^2}\right),$$

so that

$$\frac{\beta_2-1}{4} = \frac{r^3\sqrt{r-2}+\nu^2(r^2+10r-12)}{2\sqrt{r-2}\sqrt{r-3}(r^2+\nu^2)},$$

and

$$\sigma^2_{a_\mu} = \frac{a^2}{n}\cdot\frac{r^3\sqrt{r-2}+\nu^2(r^2+10r-12)}{2\sqrt{r-2}\sqrt{r-3}(r^2+\nu^2)}.$$

The efficiency of the method of moments for scaling is thus

$$\frac{\sqrt{r-2}\sqrt{r-3}\sqrt{r+4}(r^2+\nu^2)}{r+1\{r^3\sqrt{r-2}+\nu^2(r^2+10r-12)\}}; \quad \ldots\ldots\ldots \quad (4)$$

when $\nu = 0$, we have for curves of Type VII. an efficiency of scaling

$$1-\frac{12}{r\sqrt{r+1}}.$$

The efficiency of the method of moments in scaling these curves vanishes at $r = 3$, where $\beta_2$ becomes infinite; for $r = 7$, the efficiency of scaling is

$$\frac{55}{2}\cdot\frac{49+\nu^2}{1715+107\nu^2},$$

varying in value from 78·57 per cent. for the symmetrical Type VII. curve, to 25·70 per cent. when $\nu\to\infty$ and the curve to Type V.

## 10. The Efficiency of the Method of Moments in fitting the Pearsonian Curves.

The Pearsonian group of skew curves are obtained as solutions of the equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x-m)}{a+bx+cx^2}; \quad \ldots\ldots\ldots \quad (5)$$

algebraically these fall into two main classes,

$$df \propto \left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2}dx$$

and

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-\nu\tan^{-1}\frac{x}{a}}dx,$$

according as the roots of the quadratic expression in (5) are real or imaginary.

The first of these forms may be rewritten

$$df \propto \left(1-\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-r\tanh^{-1}\frac{x}{a}}\,dx,$$

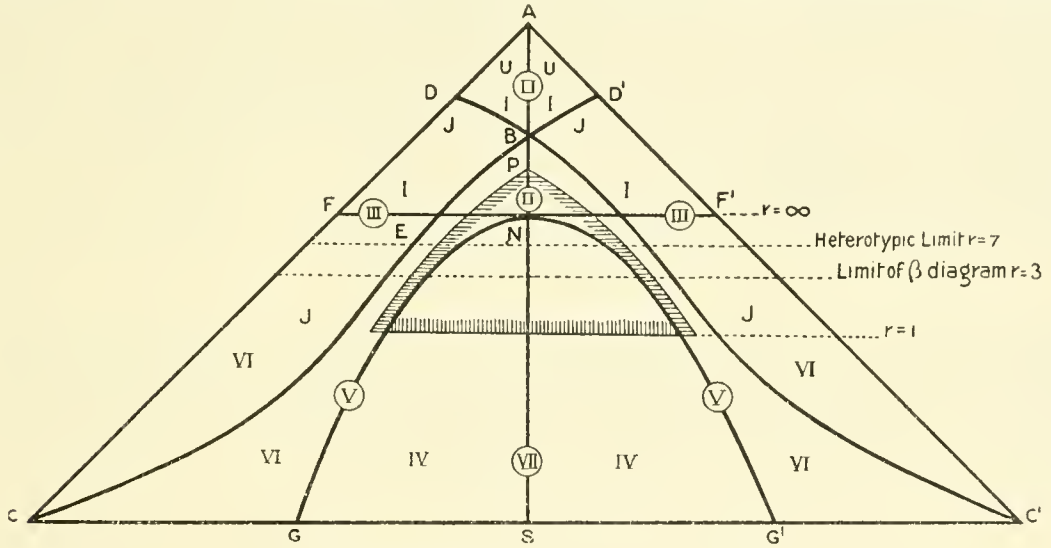
$r$ being negative, showing its affinity with the second class.

In order that these expressions may represent frequency curves, it is necessary that the integral over the whole range of the curve should be finite; this restriction acts in two ways:—

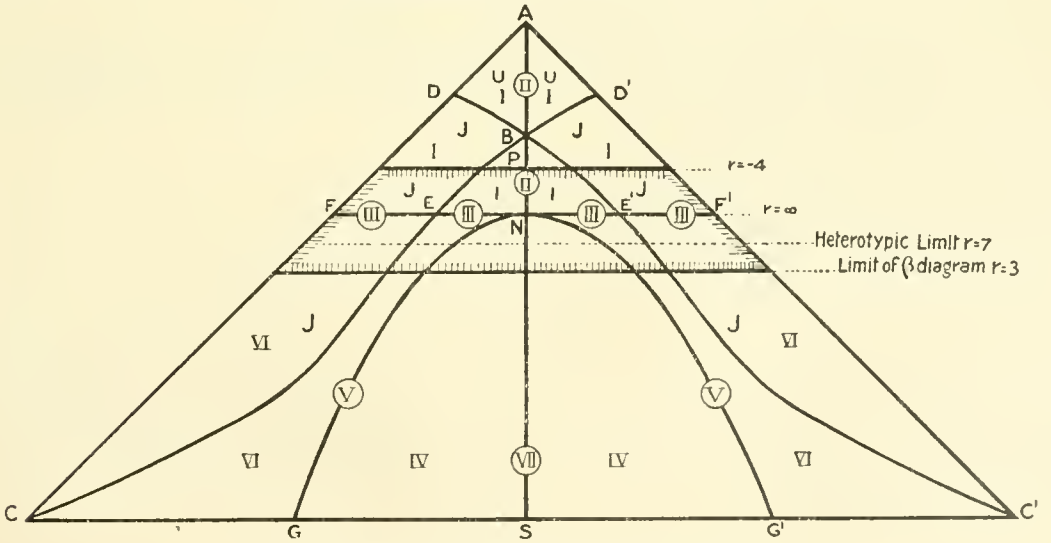
(1) When the curve terminates at a finite value of $x$, say $x = a_2$, the power to which $a_2 - x$ is raised must be greater than $-1$.

(2) When the curve extends to infinity, the ordinate, when $x$ is large, must diminish more rapidly than $\frac{1}{x}$;

In Fig. 2 is shown a conspectus of all possible frequency curves of the Pearsonian type;




A. Showing region of validity of first moment.




B. Showing region of validity of second moment.

Fig. 2. Conspectus of Pearsonian system of frequency curves.

the lines AC and AC′ represent the limits along which the area between the curve and a vertical ordinate tends to infinity, and on which $m_1$, or $m_2$, takes the value $-1$; the line CC′ represents the limit at which unbounded curves enclose an infinite area with the horizontal axis; at this limit $r = -1$.

The symmetrical curves of Type II.

$$df \propto \left(1 - \frac{x^2}{a^2}\right)^{-\frac{r+2}{2}}$$

extend from the point N, representing the normal curve, at which $r$ is infinite, through the point P at which $r = -4$, and the curve is a *parabola*, to the point B ($r = -2$), where the curve takes the form of a *rectangle*; from this point the curves are U-shaped, and at A, when the arms of U are hyperbolic, we have the limiting curve of this type, which is the discontinuous distribution of *equal or unequal dichotomy* ($r = 0$).

The unsymmetrical curves of Type I. are divided by Pearson into three classes according as the terminal ordinate is infinite at neither end, at one end (J curves), or at both ends (U curves); the dividing lines are C′BD and CBD′, along which one of the terminal ordinates are finite ($m_1$, or $m_2$, $= 0$); at the point B, as we have seen, both terminal ordinates are finite.

The same line of division divides the curves of Type III.,

$$df \propto x^p e^{-x}\, dx,$$

at the point E ($p = 0$), representing a simple exponential curve; the J curves of Type III. extend to F ($p = -1$), at which point the integral ceases to converge. In curves of Type III., $r$ is infinite; $\nu$ is also infinite, but one of the quantities $m_1$ and $m_2$ is finite, or zero ($= p$); as $p$ tends to infinity we approach the normal curve

$$df \propto e^{-\frac{1}{2}x^2}\, dx.$$

Type VI., like Type III., consists of curves bounded only at one end; here $r$ is positive, and both $m_1$ and $m_2$ are finite or zero. For the J curves of Type VI. both $m_1$ and $m_2$ are negative, but for the remainder of these curves they are of opposite sign, the negative index being the greater by at least unity in order that the representative point may fall above CC′ ($r = -1$).

Type V. is here represented by a parabola separating the regions of Types IV. and VI.; the typical equation of this type of curve is

$$df \propto x^{-\frac{r+3}{2}} e^{-\frac{1}{x}} dx.$$

As $r$ tends to infinity the curve tends to the normal form; the integral does not become divergent until $\frac{r+3}{2} = 1$, or $r = -1$. On curves of Type V., then, $r$ is finite or zero, but $\nu$ is infinite.

In Type IV.

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-\nu \tan^{-1}\frac{x}{a}};$$

we have written $\nu$, not as previously for the difference between $m_1$ and $m_2$, for these quantities are now complex, and their difference is a pure imaginary, but for the difference divided by $\sqrt{-1}$; $\nu$ is then real and finite throughout Type IV., and it vanishes along the line NS, representing the symmetrical curves of Type VII.

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}}$$

from $r = \infty$ to $r = -1$.

The Pearsonian system of frequency curves has hitherto been represented by the diagram (13, p. 66), in which the co-ordinates are $\beta_1$ and $\beta_2$. This is an unsymmetrical diagram which, since $\beta_1$ is necessarily positive, places the symmetrical curves on a boundary, whereas they are the central types from which the unsymmetrical curves diverge on either hand; further, neither of the limiting conditions of these curves can be shown on the $\beta$ diagram; the limit of the U curves is left obscure,* and the other limits are either projected to infinity, or, what is still more troublesome, the line at infinity cuts across the diagram, as occurs along the line $r = 3$, for there $\beta_2$ becomes infinite. This diagram thus excludes all curves of Types VII., IV., V., and VI., for which $r < 3$.

In the $\beta$ diagram the condition $r =$ constant yields a system of concurrent straight lines. The basis of the representation in fig. 2 lies in making these lines parallel and horizontal, so that the ordinate is a function of $r$ only. We have chosen $r = y - \frac{1}{y}$, and have represented the limiting types by the simplest geometrical forms, straight lines and parabolas, by taking

$$4\epsilon = r^2 + \nu^2 = \frac{(\overline{1+y}^2 + x^2)(\overline{1-y}^2 - x^2)}{y(x^2+y)}.$$

It might have been thought that use could have been made of the criterion,

$$\kappa_2 = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} = 1 - \frac{r^2}{4\epsilon},$$

by which PEARSON distinguishes these curves; but this criterion is only valid in the region treated by PEARSON. For when $r = 0$, $\kappa_2 = 1$, and we should have to place a variety of curves of Types VII., IV., V., and VI., all in Type V. in order to adhere to the criterion.

This diagram gives, I believe, the simplest possible *conspectus* of the whole of the Pearsonian system of curves; the inclusion of the curves beyond $r = 3$ becomes neces-

* The true limit is the line $\beta_2 = \beta_1 + 1$, along which the curves degenerate into simple dichotomies.

sary as soon as we take a view unrestricted by the method of moments; of the so-called heterotypic curves between $r = 3$ and $r = 7$ it should be noticed that they not only fall into the ordinary Pearsonian types, but have finite values for the moment coefficients $\beta_1$ and $\beta_2$; they differ from those in which $r$ exceeds 7, merely in the fact that the value of $\beta_2$, *calculated from the fourth moment of a sample*, has an infinite probable error. It is therefore evident that this is not the right method to treat the sample, but this does not constitute, as it has been called, "the failure of Type IV.," but merely the failure of the method of moments to make a valid estimate of the form of these curves. As we shall see in more detail, the method of moments, when its efficiency is tested, fails equally in other parts of the diagram.

In expression (3) we have found that the efficiency of the method of moments for location of a curve of Type IV. is

$$\mathrm{E} = \frac{r^2 \overline{r-1}\,(\overline{r+4}^2 + r^2)}{\overline{r+1}\,\overline{r+2}\,\overline{r+4}\,(r^2+r^2)},$$

whence if we substitute for $r$ and $\nu$ in terms of the co-ordinates of our diagram, we obtain a general formula for the efficiency of the method of moments in locating Pearsonian curves, which is applicable within the boundary of the zero contour (fig. 3). This may

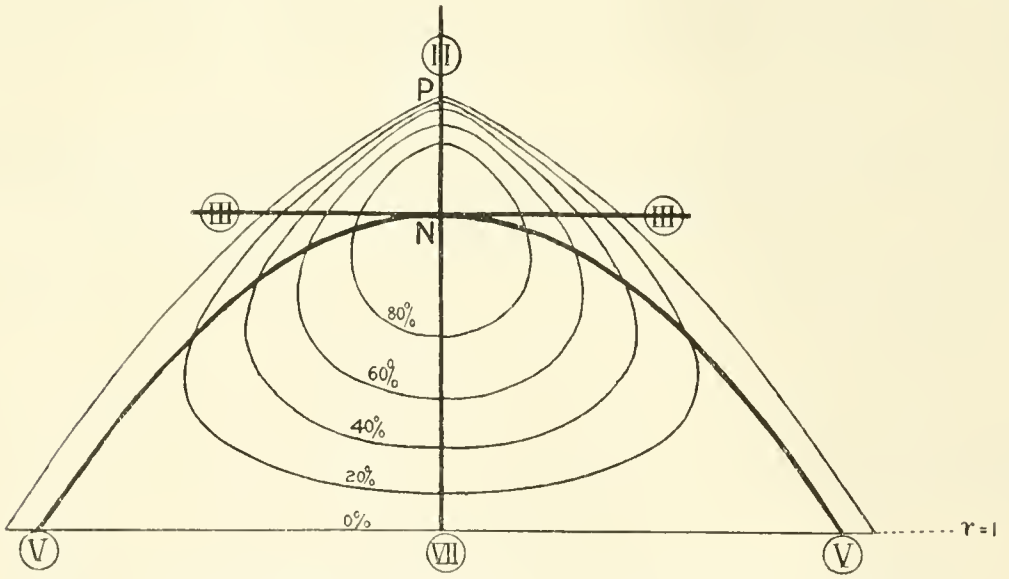



Fig. 3. Region of validity of the first moment (the mean) applied in the location of Pearsonian curves showing contours of efficiency.

be called the region of validity of the first moment; it is bounded at the base by the line $r = 1$, so that the first moment is valid far beyond the heterotypic limit; its other boundary, however, represents those curves which make a finite angle with the axis at the end of their range ($m_1$, or $m_2$, $= 1$); all J curves ($m_1$, or $m_2$, $< 0$) are thus excluded. This boundary has a double point at P, which thus forms the apex of the region of validity.

In fig. 3 are shown the contours along which the efficiency is 20, 40, 60, and 80 per cent. For high efficiencies these contours tend to the system of ellipses,

$$8x^2+6y^2 = 1-\mathrm{E}.$$

In a similar manner, we have obtained in expression (4) the efficiency of the second moment in fitting Pearsonian curves. The region of validity in this case is shown in fig. 4 ; this region is bounded by the lines $r = 3$, $r = -4$, and by the limits

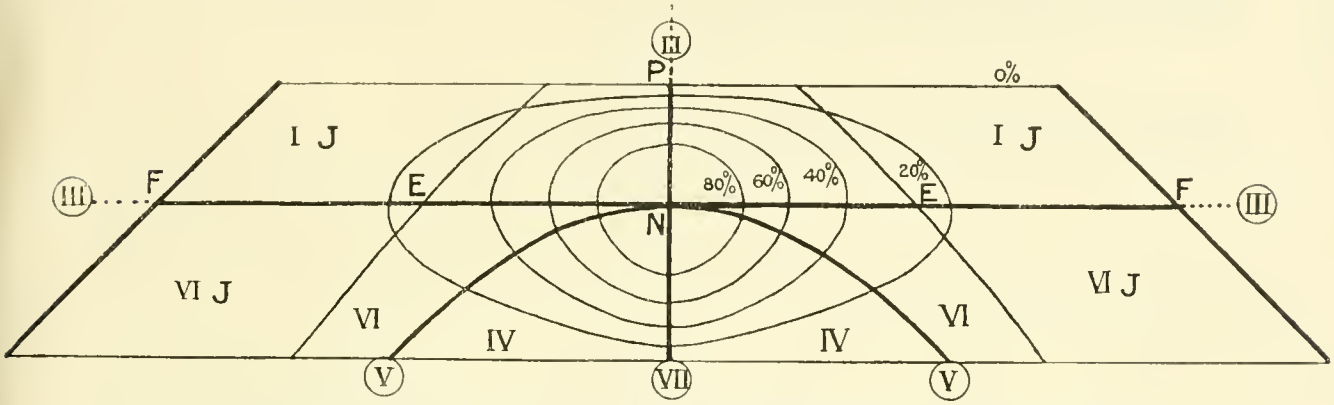



Fig. 4. Region of validity of the second moment (standard deviation) applied in scaling of Pearsonian curves, showing contours of efficiency.

($m_1$, or $m_2$, $= -1$) on which $r^2+r^2$ vanishes. This statistic is therefore valid for certain J curves, though the maximum efficiency among the J curves is about 30 per cent. As before, the contours are centred about the normal curve (N) and for high efficieneies tend to the system of concentric circles,

$$12x^2+12y^2 = 1-\mathrm{E},$$

showing that the region of high efficiency is somewhat more restricted for the second moment, as compared to the first.

The lower boundary to the efficiencies of these statistics is due merely to their probable errors becoming infinite, a weakness of the method of moments which has been partially recognised by the exclusion of the so-called heterotypic curves ($r < 7$). The stringency of the upper boundary is much more unexpected ; the probable errors of the moments do not here become infinite ; only the ratio of the probable errors of the moments to the probable error of the corresponding optimum statistics is great and tends to infinity as the size of the sample is increased.

That this failure as regards location occurs when the curve makes a finite angle with the axis may be seen by considering the occurrence of observations near the terminus of the curve.

Let

$$df = kx^a\,dx$$

in the neighbourhood of the terminus, then the chance of an observation falling within a distance $x$ of the terminus is

$$f = \frac{k}{\alpha+1} x^{\alpha+1} = k' x^{\alpha+1},$$

and the chance of $n$ observations all failing to fall in this region is

$$(1-f)^n$$

or, when $n$ is great, and $f$ correspondingly small,

$$e^{-fn}.$$

Equating this to any finite probability, $e^{-a}$, we have

$$k' x^{\alpha+1} = \frac{a}{n},$$

or, in other words, if we use the extreme observation as a means of locating the terminus, the error, $x$, is proportional to

$$n^{-\frac{1}{\alpha+1}};$$

when $\alpha < 1$, this quantity diminishes more rapidly than $n^{-\frac{1}{2}}$, and consequently for large samples it is much more accurate to locate the curve by the extreme observation than by the mean.

Since it might be doubted whether such a simple method could really be more accurate than the process of finding the actual mean, we will take as example the location of the curve (B) in the form of a rectangle,

$$df = \frac{dx}{a}, \qquad m - \frac{a}{2} < x < m + \frac{a}{2},$$

and

$$df = 0,$$

outside these limits.

This is one of the simplest types of distribution, and we may readily obtain examples of it from mathematical tables. The mean of the distribution is $m$, and the standard deviation $\frac{a}{\sqrt{12}}$, the error $m_\mu - m$, of the mean obtained from $n$ observations, when $n$ is reasonably large, is therefore distributed according to the formula

$$\frac{1}{a}\sqrt{\frac{6n}{\pi}}\, e^{-\frac{6nx^2}{a^2}}\, dx.$$

The difference of the extreme observation from the end of the range is distributed according to the formula

$$\frac{n}{a} e^{-\frac{n\xi}{a}}\, d\xi;$$

if $\xi$ is the difference at one end of the range and $\eta$ the difference at the other end, the joint distribution (since, when $n$ is considerable, these two quantities may be regarded as independent) is

$$\frac{n^2}{a^2} e^{-\frac{n}{a}(\xi+\eta)}\, d\xi\, d\eta.$$

Now if we take the mean of the extreme observations of the sample, our error is

$$\tfrac{1}{2}(\xi - \eta),$$

for which we write $x$; writing also $y$ for $\xi + \eta$, we have the joint distribution of $x$ and $y$,

$$\frac{n^2}{a^2} e^{-\frac{n}{a}y}\, dx\, dy.$$

For a given value of $x$ the values of $y$ range from $2|x|$ to $\infty$, whence, integrating with respect to $y$, we find the distribution of $x$ to be

$$df = \frac{n}{a} e^{-\frac{2n}{a}|x|}\, dx,$$

the double exponential curve shown in fig. 5.

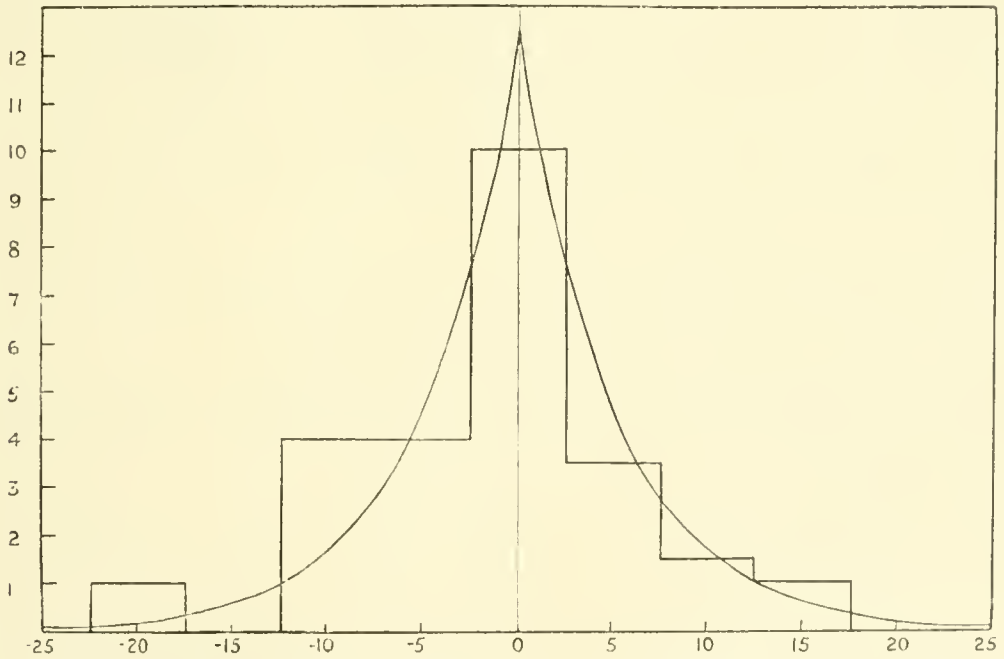



Fig. 5. Double exponential frequency curve, showing distribution of 25 deviations.

The two error curves are thus of a radically different form, and strictly no value for the efficiency can be calculated; if, however, we consider the ratio of the two standard deviations, then

$$\frac{\sigma_{\hat{m}}^2}{\sigma_{m_\mu}^2} = \frac{a^2}{2n^2} \div \frac{a^2}{12n} = \frac{6}{n}$$

when $n$ is large, a quantity which diminishes indefinitely as the sample is increased.

For example, we have taken from Vega (14) sets of digits from the table of Natural Logarithms to 48 places of decimals. The last block of four digits was taken from the logarithms of 100 consecutive numbers from 101 to 200, giving a sample of 100 numbers distributed evenly over a limited range. It is sufficient to take the three first digits to the nearest integer; then each number has an equal chance of all values between 0 and 1000. The true mean of the population is 500, and the standard deviation 289. The standard error of the mean of a sample of 100 is therefore 28·9.

Twenty-five such samples were taken, using the last five blocks of digits, for the logarithms of numbers from 101 to 600, and the mean determined merely from the highest and lowest number occurring, the following values were obtained :—

| Digits. | 1st hundred. | | | 2nd hundred. | | | 3rd hundred. | | | 4th hundred. | | | 5th hundred. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lowest. | Highest. | $\hat{m}-m$. | Lowest. | Highest. | $\hat{m}-m$. | Lowest. | Highest. | $\hat{m}-m$. | Lowest. | Highest. | $\hat{m}-m$. | Lowest. | Highest. | $\hat{m}-m$. |
| 45–48 | 24 | 978 | + 1·0 | 39 | 980 | + 9·5 | 1 | 999 | 0 | 16 | 983 | − 0·5 | 18 | 994 | +6·0 |
| 41–41 | 35·5 | 993 | +14·0 | 3 | 960 | −18·5 | 6 | 997 | +1·5 | 1 | 978 | −10·5 | 4 | 979 | −8·5 |
| 37–40 | 9 | 988 | − 1·5 | 11 | 999 | + 5·0 | 31 | 984 | +7·5 | 4 | 978 | − 9·0 | 2 | 986 | −6·0 |
| 33–36 | 7 | 995 | + 1·0 | 13 | 997 | + 5·0 | 4 | 998 | +1·0 | 0 | 994 | − 3·0 | 3 | 981 | −8·0 |
| 29–32 | 1 | 988 | − 5·5 | 3 | 988 | − 4·5 | 4 | 992 | −2·0 | 1 | 996 | − 1·5 | 21 | 977 | −1·0 |

It will be seen that these errors rarely exceed one-half of the standard error of the mean of the sample. The actual mean square error of these 25 values is 6·86, while the calculated value, $\sqrt{50}$, is 7·07. It will therefore be seen that, with samples of only 100, there is no exaggeration in placing the efficiency of the method of moments as low as 6 per cent. in comparison with the more accurate method, which in this case happens to be far less laborious.

Such a value for the efficiency of the mean in this case is, however, purely conventional, since the curve of distribution is outside the region of its valid application, and the two curves of sampling do not tend to assume the same form. It is, however, convenient to have an estimate of the effectiveness of statistics for small samples, and in such cases we should prefer to treat the curve of distribution of the statistic as an error curve, and to judge the effectiveness of the statistic by the *intrinsic accuracy* of the curve as defined in Section 9. Thus the intrinsic accuracy of the curve of distribution of the mean of all the observations is

$$\frac{12n}{a^2}.$$

while that of the mean of the extreme values is

$$\frac{4n^2}{a^2},$$

so yielding a ratio $3/n$. It is probable that this quantity may prove a suitable substitute for the efficiency of a statistic for curves beyond its region of validity.

To determine the efficiency of the moment coefficients $\beta_1$ and $\beta_2$ in determining the form of a Pearsonian curve, we must in general apply the method of Section 8 to the calculation of the simultaneous distribution of the four parameters of those curves when estimated by the method of maximum likelihood. Expressing the curve by the formula appropriate to Type IV., we are led to the determinant

$$\begin{vmatrix}
\frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{a^2(\overline{r+4}^2+\nu^2)} & -\frac{\overline{r+1}\,\overline{r+2}\,\nu}{a^2(\overline{r+4}^2+\nu^2)} & \frac{\overline{r+1}\,\overline{r+2}}{a(\overline{r+2}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} \\
-\frac{\overline{r+1}\,\overline{r+2}\,\nu}{a^2(\overline{r+4}^2+\nu^2)} & \frac{\overline{r+1}\,(2\overline{r+4}+\nu^2)}{a^2(\overline{r+4}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{r+2+\nu^2}{a(\overline{r+2}^2+\nu^2)} \\
\frac{\overline{r+1}\,\overline{r+2}}{a(\overline{r+2}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{\partial^2}{\partial\nu^2}\log \mathrm{F} & \frac{\partial^2}{\partial\nu\,\partial r}\log \mathrm{F} \\
-\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{r+2+\nu^2}{a(\overline{r+2}^2+\nu^2)} & \frac{\partial^2}{\partial\nu\,\partial r}\log \mathrm{F} & \frac{\partial^2}{\partial r^2}\log \mathrm{F}
\end{vmatrix}$$

as the Hessian of —L, when

$$\mathrm{F} = e^{-\frac{1}{2}\nu\pi}\int_0^\pi e^{\nu\theta}\sin^r\theta\,d\theta.$$

The ratios of the minors of this determinant to the value of the determinant give the standard deviations and correlations of the optimum values of the four parameters obtained from a number of large samples.

In discussing the efficiency of the method of moments in respect of the *form* of the curve, it is doubtful if it be possible to isolate in a unique and natural manner, as we have done in respect of *location* and *scaling*, a series of parameters which shall successively represent different aspects of the process of curve fitting. Thus we might find the efficiencies with which $r$ and $\nu$ are determined by the method of moments, or those of the parametric functions corresponding to $\beta_1$ and $\beta_2$, or we might use $m_1$ and $m_2$ as independent parameters of form; but in all these cases we should be employing an arbitrary pair of measures to indicate the relative magnitude of corresponding contour ellipses of the two frequency surfaces.

For the symmetrical series of curves, the Types II. and VII., the two systems of

ellipses are coaxial, the deviations of $r$ and $\nu$ being uncorrelated ; in the case of Type VII. we put $\nu = 0$, in the determinant given above, which then becomes

$$\begin{vmatrix} \frac{\sqrt{r+1}\,\sqrt{r+2}}{r+4} & 0 & \frac{r+1}{r+2} & 0 \\ 0 & \frac{2\sqrt{r+1}}{r+4} & 0 & \frac{1}{r+2} \\ \frac{r+1}{r+2} & 0 & \frac{1}{2}\mathrm{F}\left(\frac{r}{2}\right) & 0 \\ 0 & \frac{1}{r+2} & 0 & \frac{1}{4}\left\{\mathrm{F}\left(\frac{r-1}{2}\right)-\mathrm{F}\left(\frac{r}{2}\right)\right\} \end{vmatrix}$$

and falls in the two factors

$$\left[\frac{r+1}{2\sqrt{r+4}}\left\{\mathrm{F}\left(\frac{r-1}{2}\right)-\mathrm{F}\left(\frac{r}{2}\right)\right\}-\frac{1}{\sqrt{r-2}^2}\right]\left[\frac{\sqrt{r+1}\,\sqrt{r+2}}{2\sqrt{r+4}}\mathrm{F}\left(\frac{r}{2}\right)-\frac{\sqrt{r+1}^2}{\sqrt{r+2}^2}\right],$$

so that

$$\mu\sigma_{\rho}^{2}=\frac{2\sqrt{r+2}^3}{\sqrt{r+2}^3\,\mathrm{F}\left(\frac{r}{2}\right)-2\sqrt{r+1}\,\sqrt{r+4}},$$

and

$$\mu\sigma_{r}^{2}=\frac{4\sqrt{r+1}\,\sqrt{r+2}^2}{\sqrt{r+1}\,\sqrt{r+2}^2\left\{\mathrm{F}\left(\frac{r-1}{2}\right)-\mathrm{F}\left(\frac{r}{2}\right)\right\}-2\sqrt{r+4}}.$$

The corresponding expressions for the method of moments are

$$\mu\sigma_{r_\mu}^{2}=\frac{3}{8}\cdot\frac{r^2\sqrt{r-2}^2(r^2+r+10)}{\sqrt{r-1}\,\sqrt{r-3}\,\sqrt{r-5}},$$

and

$$\mu\sigma_{r_\mu}^{2}=\frac{2}{3}\cdot\frac{r\sqrt{r-1}^3\sqrt{r-3}\,(r^2-r+18)}{\sqrt{r-5}\,\sqrt{r-7}}.$$

Since for moderately large values of $r$, we have, approximately,

$$\sqrt{r+2}^3\,\mathrm{F}\left(\frac{r}{2}\right)-2\sqrt{r+1}\,\sqrt{r+4}=\frac{16}{3}\left(1-\frac{1}{5\sqrt{r+2}^2}\right),$$

and

$$\sqrt{r+1}\,\sqrt{r+2}^2\left\{\mathrm{F}\left(\frac{r-1}{2}\right)-\mathrm{F}\left(\frac{r}{2}\right)\right\}-2\sqrt{r+1}\,\sqrt{r+4}=6-\frac{4}{\sqrt{r+2}^2};$$

we have, approximately, for the efficiency of $\nu_\mu$,

$$\frac{(\overline{r+2}^3+\frac{1}{5}\overline{r+2}\ldots)\overline{r-1}\,\overline{r-3}\,\overline{r-5}}{(r^2+r+10)\,r^2\overline{r-2}^2},$$

or, when $r$ is great,

$$1-\frac{28{\cdot}8}{r^2};$$

and for the efficiency of $r_\mu$,

$$\frac{(\overline{r+2}^2+\frac{2}{3}\ldots)\overline{r+1}^2\,\overline{r-5}\,\overline{r-7}}{(r^2-r+18)\,r\,\overline{r-1}^2\,\overline{r-3}};$$

or, when $r$ is great,

$$1-\frac{53{\cdot}3}{r^2}.$$

The following table gives the values of the transcendental quantities required, and the efficiency of the method of moments in estimating the value of $\nu$ and $r$ from samples drawn from Type VII. distribution.

| $r$. | $\overline{r+2}^2\,\mathsf{F}\left(\frac{r}{2}\right)$ $-2\overline{r+1}\,r+4.$ | Efficiency of $\nu_\mu$. | $\overline{r+1}^2\,\overline{r+2}^2$ $\times\left\{\mathsf{F}\left(\frac{r-1}{2}\right)-\mathsf{F}\left(\frac{r}{2}\right)\right\}$ $-2\overline{r+1}\,r+4.$ | Efficiency of $r_\mu$. |
|---|---|---|---|---|
| 5 | 5·31271 | 0 | | |
| 6 | 5·31736 | 0·2572 | | |
| 7 | 5·32060 | 0·4338 | 5·9473 | 0 |
| 8 | 5·32296 | 0·5569 | 5·9574 | 0·1687 |
| 9 | 5·32472 | 0·6449 | 5·9649 | 0·3130 |
| 10 | 5·32607 | 0·7097 | 5·9706 | 0·4403 |
| 11 | 5·32713 | 0·7586 | 5·9750 | 0·5207 |
| 12 | 5·32797 | 0·7963 | 5·9787 | 0·5935 |
| 13 | 5·32866 | 0·8259 | 5·9810 | 0·6519 |
| 14 | 5·32919 | 0·8497 | 5·9839 | 0·6990 |
| 15 | | | 5·9853 | 0·7376 |
| 16 | | | 5·9870 | 0·7694 |
| 17 | | | 5·9883 | 0·7959 |
| 18 | | | 5·9895 | 0·8182 |

It will be seen that we do not attain to 80 per cent. efficiency in estimating the form of the curve until $r$ is about 17·2, which corresponds to $\beta_2 = 3{\cdot}42$. Even for symmetrical curves higher values of $\beta_2$ imply that the method of moments makes use of less than four-fifths of the information supplied by the sample.

On the other side of the normal point, among the Type II. curves, very similar formulæ apply. The fundamental Hessian is

$$\begin{vmatrix} \frac{\sqrt{r-1}\,\sqrt{r-2}}{r-4} & 0 & -\frac{r-1}{r-2} & 0 \\ 0 & \frac{2\sqrt{r-1}}{r-4} & 0 & -\frac{1}{r-2} \\ -\frac{r-1}{r-2} & 0 & \frac{1}{2}\mathrm{F}\left(\frac{r-2}{2}\right) & 0 \\ 0 & -\frac{1}{r-2} & 0 & \frac{1}{4}\left\{\mathrm{F}\left(\frac{r-2}{2}\right)-\mathrm{F}\left(\frac{r-1}{2}\right)\right\} \end{vmatrix}$$

where $r$ is written for the positive quantity. $-r$, whence

$$n\sigma_\phi^2 = \frac{2\sqrt{r-2}^3}{\sqrt{r-2}^3\,\mathrm{F}\left(\frac{r-2}{2}\right)-2\sqrt{r-1}\,\sqrt{r-4}}$$

and

$$n\sigma_r^2 = \frac{4\sqrt{r-1}\,\sqrt{r-2}^2}{\sqrt{r-1}\,\sqrt{r-2}^2\left\{\mathrm{F}\left(\frac{r-2}{2}\right)-\mathrm{F}\left(\frac{r-1}{2}\right)\right\}-2\sqrt{r-4}}$$

Now since

$$\mathrm{F}\left(\frac{r-2}{2}\right) = \mathrm{F}\left(\frac{r-4}{2}\right) - \frac{4}{\sqrt{r-2}^2},$$

it follows that

$$\sqrt{r-2}^3\,\mathrm{F}\left(\frac{r-2}{2}\right)-2\sqrt{r-1}\,\sqrt{r-4} = \sqrt{r-2}^3\,\mathrm{F}\left(\frac{r-4}{2}\right)-2r\sqrt{r-3},$$

which is the same function of $r-4$ as

$$\sqrt{r+2}^3\,\mathrm{F}\left(\frac{r}{2}\right)-2\sqrt{r+1}\,\sqrt{r+4}$$

is of $r$.

In a similar manner

$$\sqrt{r+1}^2\,\sqrt{r-2}^2\left\{\mathrm{F}\left(\frac{r-2}{2}\right)-\mathrm{F}\left(\frac{r-1}{2}\right)\right\}-2\sqrt{r-1}\,\sqrt{r-4}$$

$$= \sqrt{r-1}^2\,\sqrt{r-2}^2\left\{\mathrm{F}\,\frac{r-4}{2}-\mathrm{F}\,\frac{r-3}{2}\right\}-2\sqrt{r-2}\,\sqrt{r+1},$$

which is the same function of $r-3$ as

$$\sqrt{r+1}^2\,\sqrt{r+2}^2\left\{\mathrm{F}\left(\frac{r-1}{2}\right)-\mathrm{F}\left(\frac{r}{2}\right)\right\}-2\sqrt{r+1}\,\sqrt{r+4}$$

is of $r$.

In all these functions and those of the following table, $r$ must be substituted as a positive quantity, although it must not be forgotten that $r$ changes sign as we pass from Type VII. to Type II., and we have hitherto adhered to the convention that $r$ is to be taken positive for Type VII. and negative for Type II.

| $r$. | $\frac{r-2^3\Gamma\left(\frac{r-2}{2}\right)}{\sqrt{2r-1}\,r-4.}$ | Efficiency of $r_{\mu_2}$. | $\frac{r-1^2\sqrt{r-2^2}}{\sqrt{2r-1}\,r-4.} \times \left\{\Gamma\left(\frac{r-2}{2}\right)-\Gamma\left(\frac{r-1}{2}\right)\right\}$ | Efficiency of $r_{\mu_2}$. |
|---|---|---|---|---|
| 2 | 4 | 0 | 4 | 0 |
| 3 | 4·93480 | 0·0576 | 5·1595 | 0·0431 |
| 4 | 5·15947 | 0·2056 | 5·5648 | 0·1445 |
| 5 | 5·23966 | 0·3590 | 5·7410 | 0·2613 |
| 6 | 5·27578 | 0·4865 | 5·8305 | 0·3708 |
| 7 | 5·29472 | 0·5857 | 5·8813 | 0·4653 |
| 8 | 5·30576 | 0·6615 | 5·9126 | 0·5441 |
| 9 | 5·31271 | 0·7198 | 5·9331 | 0·6090 |
| 10 | 5·31736 | 0·7650 | 5·9473 | 0·6624 |
| 11 | 5·32060 | 0·8005 | 5·9574 | 0·7063 |
| 12 | 5·32296 | 0·8287 | 5·9649 | 0·7427 |
| 13 | 5·32472 | 0·8516 | 5·9706 | 0·7731 |
| 14 | 5·32607 | 0·8702 | 5·9750 | 0·7986 |
| 15 | | | 5·9787 | 0·8202 |

In both cases the region of validity is bounded by the rectangle, at the point B (fig. 2, p. 343). Efficiency of 80 per cent. is reached when $r$ is about 14·1 ($\beta_2 = 2·65$). Thus for symmetrical curves of the Pearsonian type we may say that the method of moments has an efficiency of 80 per cent. or more, when $\beta_2$ lies between 2·65 and 3·42. The limits within which the values of the parameters obtained by moments cannot be greatly improved are thus much narrower than has been imagined.

## 11. The Reason for the Efficiency of the Method of Moments in a Small Region surrounding the Normal Curve.

We have seen that the method of moments applied in fitting Pearsonian curves has an efficiency exceeding 80 per cent. only in the restricted region for which $\beta_2$ lies between the limits 2·65 and 3·42, and as we have seen in Section 8, for which $\beta_1$ does not exceed 0·1. The contours of equal efficiency are nearly circular or elliptical within these limits, if the curves are represented as in fig. 2, p. 343, and are ultimately centred round the normal point, at which point the efficiencies of all parameters tend to 100 per cent. It was, of course, to be expected that the first two moments would have 100 per cent. efficiencies at this point, for they happen to be the optimum statistics for fitting the normal curve. That the moment coefficients $\beta_1$ and $\beta_2$ also tend to 100 per cent. efficiency in this region suggests that in the immediate neighbourhood of the normal

curve the departures from normality specified by the Pearsonian formula agree with those of that system of curves for which the method of moments gives the solution of maximum likelihood.

The system of curves for which the method of moments is the best method of fitting may easily be deduced, for if the frequency in the range $dx$ be

$$y\,(x, \theta_1, \theta_2, \theta_3, \theta_4)\,dx,$$

then

$$\frac{\partial}{\partial\theta}\log y$$

must involve $x$ only as polynomials up to the fourth degree; consequently

$$y = e^{-\alpha^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)},$$

the convergence of the probability integral requiring that the coefficient of $x^4$ should be negative, and the five quantities $\alpha$, $p_1$, $p_2$, $p_3$, $p_4$ being connected by a single relation, representing the fact that the total probability is unity.

Typically these curves are bimodal, and except in the neighbourhood of the normal point are of a very different character from the Pearsonian curves. Near this point, however, they may be shown to agree with the Pearsonian type; for let

$$y = Ce^{-\frac{x^2}{2\sigma^2}+k_1\frac{x^3}{\sigma^3}+k_2\frac{x^4}{\sigma^4}}$$

represent a curve of the quartic exponent, sufficiently near to the normal curve for the squares of $k_1$ and $k_2$ to be neglected, then

$$\frac{d}{dx}\log y = -\frac{x}{\sigma^2}\left(1-3k_1\frac{x}{\sigma}-4k_2\frac{x^2}{\sigma^2}\right)$$

$$= -\frac{x}{\sigma^2\left(1+3k_1\frac{x}{\sigma}+4k_2\frac{x^2}{\sigma^2}\right)},$$

neglecting powers of $k_1$ and $k_2$. Since the only terms in the denominator constitute a quadratic in $x$, the curve satisfies the fundamental equation of the Pearsonian type of curves. In the neighbourhood of the normal point, therefore, the Pearsonian curves are equivalent to curves of the quartic exponent; it is to this that the efficiency of $\mu_3$ and $\mu_4$ in the neighbourhood of the normal curve, is to be ascribed.

## 12. Discontinuous Distributions.

The applications hitherto made of the optimum statistics have been problems in which the data are ungrouped, or at least in which the grouping intervals are so small as not to disturb the values of the derived statistics. By grouping, these continuous

distributions are reduced to discontinuous distributions, and in an exact discussion must be treated as such.

If $p_s$ be the probability of an observation falling in the cell $(s)$, $p_s$ being a function of the required parameters $\theta_1, \theta_2 \ldots$; and in a sample of N, if $n_s$ are found to fall into that cell, then

$$\mathrm{S}(\log f) = \mathrm{S}(n_s \log p_s).$$

If now we write $\overline{n}_s = p_s \mathrm{N}$, we may conveniently put

$$\mathrm{L} = \mathrm{S}\left(n_s \log \frac{n_s}{\overline{n}_s}\right),$$

where L differs by a constant only from the logarithm of the likelihood, with sign reversed, and therefore the method of the optimum will consist in finding the *minimum* value of L. The equations so found are of the form

$$\frac{\partial \mathrm{L}}{\partial \theta} = -\mathrm{S}\left(\frac{n_s}{\overline{n}_s}\frac{\partial \overline{n}_s}{\partial \theta}\right) = 0. \quad \ldots \quad (6)$$

It is of interest to compare these formulæ with those obtained by making the Pearsonian $\chi^2$ a minimum.

For

$$\chi^2 = \mathrm{S}\frac{(n_s - \overline{n}_s)^2}{n_s},$$

and therefore

$$1 + \chi^2 = \mathrm{S}\left(\frac{n_s^2}{\overline{n}_s}\right),$$

so that on differentiating by $d\theta$, the condition that $\chi^2$ should be a minimum for variations of $\theta$ is

$$-\mathrm{S}\left(\frac{n_s^2}{\overline{n}_s^2}\frac{\partial \overline{n}_s}{\partial \theta}\right) = 0. \quad \ldots \quad (7)$$

Equation (7) has actually been used (12) to "improve" the values obtained by the method of moments, even in cases of normal distribution, and the Poisson series, where the method of moments gives a strictly sufficient solution. The discrepancy between these two methods arises from the fact that $\chi^2$ is itself an approximation, applicable only when $\overline{n}_s$ and $n_s$ are large, and the difference between them of a lower order of magnitude. In such cases

$$\mathrm{L} = \mathrm{S}\left(n_s \log \frac{n_s}{\overline{n}_s}\right) = \mathrm{S}\left(\overline{m+x} \log \frac{m+x}{m}\right) = \mathrm{S}\left\{x + \frac{x^2}{2m} - \frac{x^3}{6m^2} \ldots\right\},$$

and since

$$\mathrm{S}(x) = 0,$$

we have, when $x$ is in all cases small compared to $m$,

$$\mathrm{L} = \tfrac{1}{2}\mathrm{S}\left(\frac{x^2}{m}\right) = \tfrac{1}{2}\chi^2$$

as a first approximation. In those cases, therefore, when $\chi^2$ is a valid measure of the departure of the sample from expectation, it is equal to 2L; in other cases the approximation fails and L itself must be used.

The failure of equation (7) in the general problem of finding the best values for the parameters may also be seen by considering cases of fine grouping, in which the majority of observations are separated into units. For the formula in equation (6) is equivalent to

$$\mathrm{S}\left(\frac{1}{\overline{n}_s}\frac{\partial \overline{n}_s}{\partial \theta}\right)$$

where the summation is taken over all the observations, while the formula of equation (7), since it involves $n_s^2$, changes its value discontinuously, when one observation is gradually increased, at the point where it happens to coincide with a second observation.

Logically it would seem to be a necessity that that population which is chosen in fitting a hypothetical population to data should also appear the best when tested for its goodness of fit. The method of the optimum secures this agreement, and at the same time provides an extension of the process of testing goodness of fit, to those cases for which the $\chi^2$ test is invalid.

The practical value of $\chi^2$ lies in the fact that when the conditions are satisfied in order that it shall closely approximate to 2L, it is possible to give a general formula for its distribution, so that it is possible to calculate the probability, P, that in a random sample from the population considered, a worse fit should be obtained; in such cases $\chi^2$ is distributed in a curve of the Pearsonian Type III.,

$$df \propto \left(\frac{\chi^2}{2}\right)^{\frac{n-3}{2}} e^{-\frac{\chi^2}{2}} d\left(\frac{\chi^2}{2}\right)$$

or

$$df \propto \mathrm{L}^{\frac{n'-3}{2}} e^{-\mathrm{L}}\, d\mathrm{L},$$

where $n'$ is one more than the number of degrees of freedom in which the sample may differ from expectation (17).

In other cases we are at present faced with the difficulty that the distribution L requires a special investigation. This distribution will in general be discontinuous (as is that of $\chi^2$), but it is not impossible that mathematical research will reveal the existence of effective graduations for the most important groups of cases to which $\chi^2$ cannot be applied.

We shall conclude with a few illustrations of important types of discontinuous distribution.

1. *The Poisson Series.*

$$e^{-m}\left(1,\ m,\ \frac{m^2}{2!},\ \ldots,\ \frac{m^r}{r!},\ \ldots\right)$$

involves only the single parameter, and is of great importance in modern statistics. For the optimum value of $m$,

$$\mathrm{S}\left\{\frac{\partial}{\partial m}(-m+r\log m)\right\} = 0,$$

whence

$$\mathrm{S}\left(\frac{r}{\hat{m}}-1\right) = 0,$$

or

$$\hat{m} = \bar{r}.$$

The most likely value of $m$ is therefore found by taking the first moment of the series.

Differentiating a second time,

$$-\frac{1}{\sigma_{\hat{m}}^2} = \mathrm{S}\left(-\frac{r}{m^2}\right) = -\frac{n}{m},$$

so that

$$\sigma_{\hat{m}}^2 = \frac{m}{n},$$

as is well known.

2. *Grouped Normal Data.*

In the case of the normal curve of distribution it is evident that the second moment is a sufficient statistic for estimating the standard deviation ; in investigating a sufficient solution for grouped normal data, we are therefore in reality finding the optimum correction for grouping ; the SHEPPARD correction having been proved only to satisfy the criterion of consistency.

For grouped normal data we have

$$p_s = \frac{1}{\sigma\sqrt{2\pi}}\int_{x_s}^{x_{s+1}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx,$$

and the optimum values of $m$ and $\sigma$ are obtained from the equations,

$$\frac{\partial \mathrm{L}}{\partial m} = \mathrm{S}\left(\frac{n_s}{p_s}\frac{\partial p_s}{\partial m}\right) = 0,$$

$$\frac{\partial \mathrm{L}}{\partial \sigma} = \mathrm{S}\left(\frac{n_s}{p_s}\frac{\partial p_s}{\partial \sigma}\right) = 0\ ;$$

or, if we write,

$$z = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

we have the two conditions,

$$S\left(\frac{n_s}{p_s}\overline{z_s - z_{s+1}}\right) = 0$$

and

$$S\left\{\frac{n_s}{p_s}\left(\frac{x_s}{\sigma}z_s - \frac{x_{s+1}}{\sigma}z_{s+1}\right)\right\} = 0.$$

As a simple example we shall take the case chosen by K. Smith in her investigation of the variation of $\chi^2$ in the neighbourhood of the moment solution (12).

Three hundred errors in right ascension are grouped in nine classes, positive and negative errors being thrown together as shown in the following table :—

| 0″·1 arc | 0–1 | 1–2 | 2–3 | 3–4 | 4–5 | 5–6 | 6–7 | 7–8 | 8–9 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency . . | 114 | 84 | 53 | 24 | 14 | 6 | 3 | 1 | 1 |

The second moment, without correction, yields the value

$$\sigma_\nu = 2{\cdot}282542.$$

Using Sheppard's correction, we have

$$\sigma_\mu = 2{\cdot}264214,$$

while the value obtained by making $\chi^2$ a minimum is

$$\sigma_{\chi^2} = 2{\cdot}355860.$$

If the latter value were accepted we should have to conclude that Sheppard's correction, even when it is small, and applied to normal data, might be altogether of the wrong magnitude, and even in the wrong direction. In order to obtain the optimum value of $\sigma$, we tabulate the values of $\frac{\partial L}{\partial \sigma}$ in the region under consideration; this may be done without great labour if values of $\sigma$ be chosen suitable for the direct application of the table of the probability integral (13, Table II.). We then have the following values :—

| $\frac{1}{\sigma}$ | 0·43 | 0·44 | 0·45 | 0·46 |
|---|---|---|---|---|
| $\frac{\partial L}{\partial \sigma}$ | +15·135 | +2·149 | −11·098 | −24·605 |
| $\Delta^2\frac{\partial L}{\partial \sigma}$ | | −0·261 | − 0·260 | |

By interpolation,

$$\frac{1}{\hat{\sigma}} = 0{\cdot}441624$$

$$\hat{\sigma} = 2{\cdot}26437.$$

We may therefore summarise these results as follows :—

| | |
|---|---|
| Uncorrected estimate of $\sigma$ . . . . . | 2·28254 |
| SHEPPARD'S correction . . . . . | −0·01833 |
| Correction for maximum likelihood . . . . . | −0·01817 |
| " Correction " for minimum $\chi^2$ . . . . . | +0·07332 |

Far from shaking our faith, therefore, in the adequacy of SHEPPARD'S correction, when small, for normal data, this example provides a striking instance of its effectiveness, while the approximate nature of the $\chi^2$ test renders it unsuitable for improving a method which is already very accurate.

It will be useful before leaving the subject of grouped normal data to calculate the actual loss of efficiency caused by grouping, and the additional loss due to the small discrepancy between moments with SHEPPARD'S correction and the optimum solution.

To calculate the loss of efficiency involved in the process of grouping normal data, let

$$v = \frac{1}{a}\int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} f(\xi)\, d\xi,$$

when $a_\sigma$ is the group interval, then

$$v = f(\xi) + \frac{a^2}{24}f''(\xi) + \frac{a^4}{1920}f^{\text{iv}}(\xi) + \frac{a^6}{322{,}560}f^{\text{vi}}(\xi) + \dots$$

$$= f(\xi)\left\{1 + \frac{a^2}{24}(\xi^2-1) + \frac{a^4}{1920}(\xi^4-6\xi^2+3) + \frac{a^6}{322{,}560}(\xi^6-15\xi^4+45\xi^2-15) + \dots\right\},$$

whence

$$\log v = \log f + \frac{a^2}{24}(\xi^2-1) - \frac{a^4}{2880}(\xi^4+4\xi^2-2) + \frac{a^6}{181{,}440}(\xi^6+6\xi^4+3\xi^2-1) - \dots,$$

and

$$\frac{\partial^2}{\partial m^2}\log v = -\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\left\{\frac{a^2}{12} - \frac{a^4}{720}(3\xi^2+2) + \frac{a^6}{30{,}240}(5\xi^4+12\xi^2+1) - \dots\right\},$$

of which the mean value is

$$-\frac{1}{\sigma^2}\left\{1 - \frac{a^2}{12} + \frac{a^4}{144} - \frac{a^6}{4320}\dots\right\},$$

neglecting the periodic terms; and consequently

$$\sigma_{\hat{m}}^2 = \frac{\sigma^2}{n}\left(1 + \frac{a^2}{12} - \frac{a^6}{2880}\cdots\right).$$

Now for the mean of ungrouped data

$$\sigma_{\hat{m}}^2 = \frac{\sigma^2}{n},$$

so that the loss of efficiency due to grouping is nearly $\frac{a^2}{12}$.

The further loss caused by using the mean of the grouped data is very small, for

$$\sigma_{\nu_1}^2 = \frac{\nu_2}{n} = \frac{\sigma^2}{n}\left(1 + \frac{a^2}{12}\right),$$

neglecting the periodic terms; the loss of efficiency by using $\nu_1$ therefore is only

$$\frac{a^6}{2880}.$$

Similarly for the efficiency for scaling,

$$\frac{\partial^2}{\partial\sigma^2}\log r = \frac{1}{\sigma^2} - \frac{3\xi^2}{\sigma^2} + \frac{1}{\sigma^2}\left\{\frac{a^2}{12}(10\xi^2-3) - \frac{a^4}{360}(9\xi^4+21\xi^2-5)\right.$$
$$\left. + \frac{a^6}{30{,}240}(26\xi^6+110\xi^4+36\xi^2-7) - \frac{a^8}{1{,}814{,}400}(51\xi^8+315\xi^6+351\xi^4-55\xi^2+9) + \ldots\right\},$$

of which the mean value is

$$-\frac{2}{\sigma^2}\left\{1 - \frac{a^2}{6} + \frac{a^4}{40} - \frac{a^6}{270} + \frac{83a^8}{129{,}600}\cdots\right\},$$

neglecting the periodic terms; and consequently

$$\sigma_{\hat{\sigma}}^2 = \frac{\sigma^2}{2n}\left\{1 + \frac{a^2}{6} + \frac{a^4}{360} - \frac{a^8}{10{,}800}\cdots\right\}.$$

For ungrouped data

$$\sigma_{\hat{\sigma}}^2 = \frac{\sigma^2}{2n},$$

so that the loss of efficiency in scaling due to grouping is nearly $\frac{a^2}{6}$. This may be made as low as 1 per cent. by keeping $a$ less than $\frac{1}{4}$.

The further loss of efficiency produced by using the grouped second moment with SHEPPARD'S correction is again very small, for

$$\sigma_{\nu_2}^2 = \frac{\nu_4 - \nu_2^2}{n} = \frac{2\sigma^4}{n}\left(1 + \frac{a^2}{6} + \frac{a^4}{360}\right)$$

neglecting the periodic terms.

Whence it appears that the further loss of efficiency is only

$$\frac{a^8}{10{,}800}.$$

We may conclude, therefore, that the high agreement between the optimum value of $\sigma$ and that obtained by SHEPPARD'S correction in the above example is characteristic of grouped normal data. The method of moments with SHEPPARD'S correction is highly efficient in treating such material, the gain in efficiency obtainable by increasing the likelihood to its maximum value is trifling, and far less than can usually be gained by using finer groups. The loss of efficiency involved in grouping may be kept below 1 per cent. by making the group interval less than one-quarter of the standard deviation.

Although for the normal curve the loss of efficiency due to moderate grouping is very small, such is not the case with curves making a finite angle with the axis, or having at an extreme a finite or infinitely great ordinate. In such cases even moderate grouping may result in throwing away the greater part of the information which the sample provides.

### 3. *Distribution of Observations in a Dilution Series.*

An important type of discontinuous distribution occurs in the application of the dilution method to the estimation of the number of micro-organisms in a sample of water or of soil. The method here presented was originally developed in connection with Mr. CUTLER'S extensive counts of soil protozoa carried out in the protozoological laboratory at Rothamsted, and although the method is of very wide application, this particular investigation affords an admirable example of the statistical principles involved.

In principle the method consists in making a series of dilutions of the soil sample, and determining the presence or absence of each type of protozoa in a cubic centimetre of the dilution, after incubation in a nutrient medium.

The series in use proceeds by powers of 2, so that the frequency of protozoa in each dilution is one-half that in the last.

The frequency at any stage of the process may then be represented by

$$m = \frac{\mu}{2^x},$$

when $x$ indicates the number of dilutions.

Under conditions of random sampling, the chance of any plate receiving 0, 1, 2, 3 protozoa of a given species is given by the Poisson series

$$e^{-m}\left(1,\ m,\ \frac{m^2}{2!},\ \frac{m^3}{3!},\ \ldots\right),$$

and in consequence the proportion of sterile plates is

$$p = e^{-m},$$

and of fertile plates

$$q = 1 - e^{-m}.$$

In general we may consider a dilution series with dilution factor $a$ so that

$$\log p = -\frac{n}{a^x},$$

and assume that $s$ plates are poured from each dilution.

The object of the method being to estimate the number $n$ from a record of the sterile and fertile plates, we have

$$\mathrm{L} = \mathrm{S}_1(\log p) + \mathrm{S}_2(\log q)$$

when $\mathrm{S}_1$ stands for summation over the sterile plates, and $\mathrm{S}_2$ for summation over those which are fertile.

Now

$$\frac{\partial p}{\partial \log n} = -\frac{\partial q}{\partial \log n} = p \log p,$$

so that the optimum value of $n$ is obtained from the equation,

$$\frac{\partial \mathrm{L}}{\partial \log n} = \mathrm{S}_1(\log p) - \mathrm{S}_2\left(\frac{p}{q}\log p\right) = 0.$$

Differentiating a second time,

$$\frac{\partial^2 \mathrm{L}}{\partial (\log n)^2} = \mathrm{S}_1(\log p) - \mathrm{S}_2\left\{\frac{p \log p}{q}\left(\log p + 1 + \frac{p \log p}{q}\right)\right\};$$

now the mean number of sterile plates is $ps$, and of fertile plates $qs$, so that the mean value of $\frac{\partial^2 \mathrm{L}}{\partial (\log n)^2}$ is

$$-\frac{1}{\sigma^2_{\log n}} = s\mathrm{S}\left\{p \log p - p \log p\left(\log p + 1 + \frac{p}{q}\log p\right)\right\} = -s\mathrm{S}\left\{\frac{p}{q}(\log p)^2\right\},$$

the summation, S, being extended over all the dilutions.

It thus appears that each plate observed adds to the weight of the determination of $\log n$ a quantity

$$w = \frac{p}{q}(\log p)^2.$$

We give below a table of the values of $p$, and of $w$, for the dilution series $\log p = 2^{-x}$ from $x = -4$ to $x = 11$.

| $x$. | $p$. | $w$. | S ($w$) (per cent.). |
|---|---|---|---|
| −4 | 0·00000014167 | 0·000036 | 0·002 |
| −3 | 0·0003354626 | 0·021477 | 0·906 |
| −2 | 0·01831564 | 0·298518 | 13·485 |
| −1 | 0·1353353 | 0·626071 | 39·865 |
| 0 | 0·3678794 | 0·581977 | 64·388 |
| 1 | 0·6065306 | 0·385355 | 80·625 |
| 2 | 0·7788009 | 0·220051 | 89·897 |
| 3 | 0·8824969 | 0·117350 | 94·842 |
| 4 | 0·9394110 | 0·060565 | 97·394 |
| 5 | 0·9692333 | 0·030764 | 98·690 |
| 6 | 0·9844965 | 0·015503 | 99·343 |
| 7 | 0·9922179 | 0·007782 | 99·671 |
| 8 | 0·9961014 | 0·003899 | 99·836 |
| 9 | 0·9980488 | 0·001951 | 99·918 |
| 10 | 0·9990239 | 0·000976 | 99·959 |
| 11 | 0·9995119 | 0·000488 | 99·979 |
| Remainder . . . . . | | 0·000488 | |
| Total . . . . . | | 2·373251 | |

For the same dilution constant the total S ($w$) is nearly independent of the particular series chosen. Its average value being $\frac{\pi^2}{6 \log \alpha}$, or in this case 2·373138. The fourth column shows the total weight attained at any stage, expressed as a percentage of that obtained from an infinite series of dilutions. It will be seen that a set of eight dilutions comprise all but about 2 per cent. of the weight. With a loss of efficiency of only 2 to $2\frac{1}{2}$ per cent., therefore, the number of dilutions which give information as to a particular species may be confined to eight. To this number must be added a number depending on the range which it is desired to explore. Thus to explore a range from 100 to 100,000 per gramme (about 10 octaves) we should require 10 more dilutions, making 18 in all, while to explore a range of a millionfold, or about 20 octaves, 28 dilutions would be needed.

In practice it would be exceedingly laborious to calculate the optimum value of $n$ for each series observed (of which 38 are made daily). On the advice of the statistical department, therefore, Mr. CUTLER adopted the plan of counting the total number of sterile plates, and taking the value of $n$ which on the average would give that number. When a sufficient number of dilutions are made, $\log n$ is diminished by $\frac{1}{s} \log \alpha$ for each additional sterile plate, and even near the ends of the series the appropriate values of $n$ may easily be tabulated. Since this method of estimation is of wide application, and appears at first sight to be a very rough one, it is important to calculate its efficiency.

For any dilution the variance in the number of sterile plates is

$$spq,$$

and as the several dilutions represent independent samples, the total variance is

$$s\mathrm{S}(pq),$$

hence

$$\sigma^2_{\log n} = \frac{(\log a)^2}{s}\mathrm{S}(pq).$$

Now $\mathrm{S}(pq)$ has an average value $\dfrac{\log 2}{\log a}$, therefore taking $a = 2$,

$$(\log a)^2 = \cdot 480453,$$

and

$$\mathrm{S}(pq) = 1$$

being very nearly constant and within a small fraction of unity; whence the efficiency of the method of counting the sterile plates is

$$\frac{6}{\pi^2 \log 2} = 87{\cdot}71 \text{ per cent.},$$

a remarkably high efficiency, considering the simplicity of the method, the efficiency being independent of the dilution ratio.

## 13. Summary.

During the rapid development of practical statistics in the past few decades, the theoretical foundations of the subject have been involved in great obscurity. Adequate distinction has seldom been drawn between the sample recorded and the hypothetical population from which it is regarded as drawn. This obscurity is centred in the so-called "inverse" methods.

On the bases that the purpose of the statistical reduction of data is to obtain statistics which shall contain as much as possible, ideally the whole, of the relevant information contained in the sample, and that the function of Theoretical Statistics is to show how such adequate statistics may be calculated, and how much and of what kind is the information contained in them, an attempt is made to formulate distinctly the types of problems which arise in statistical practice.

Of these, problems of Specification are found to be dominated by considerations which may change rapidly during the progress of Statistical Science. In problems of Distribution relatively little progress has hitherto been made, these problems still affording a field for valuable enquiry by highly trained mathematicians. The principal purpose of this paper is to put forward a general solution of problems of Estimation.

Of the criteria used in problems of Estimation only the criterion of Consistency has hitherto been widely applied; in Section 5 are given examples of the adequate and inadequate application of this criterion. The criterion of Efficiency is shown to be a special but important case of the criterion of Sufficiency, which latter requires that the whole of the relevant information supplied by a sample shall be contained in the statistics calculated.

In order to make clear the nature of the general method of satisfying the criterion of Sufficiency, which is here put forward, it has been thought necessary to reconsider BAYES' problem in the light of the more recent criticisms to which the idea of "inverse probability" has been exposed. The conclusion is drawn that two radically distinct concepts, both of importance in influencing our judgment, have been confused under the single name of *probability*. It is proposed to use the term *likelihood* to designate the state of our information with respect to the parameters of hypothetical populations, and it is shown that the quantitative measure of likelihood does not obey the mathematical laws of probability.

A proof is given in Section 7 that the criterion of Sufficiency is satisfied by that set of values for the parameters of which the likelihood is a maximum, and that the same function may be used to calculate the efficiency of any other statistics, or, in other words, the percentage of the total available information which is made use of by such statistics.

This quantitative treatment of the information supplied by a sample is illustrated by an investigation of the efficiency of the method of moments in fitting the Pearsonian curves of Type III.

Section 9 treats of the location and scaling of Error Curves in general, and contains definitions and illustrations of the *intrinsic accuracy*, and of the *centre of location* of such curves.

In Section 10 the efficiency of the method of moments in fitting the general Pearsonian curves is tested and discussed. High efficiency is only found in the neighbourhood of the normal point. The two causes of failure of the method of moments in locating these curves are discussed and illustrated. The special cause is discovered for the high efficiency of the third and fourth moments in the neighbourhood of the normal point.

It is to be understood that the low efficiency of the moments of a sample in estimating the form of these curves does not at all diminish the value of the notation of moments as a means of the comparative specification of the form of such curves as have finite moment coefficients.

Section 12 illustrates the application of the method of maximum likelihood to discontinuous distributions. The POISSON series is shown to be sufficiently fitted by the mean. In the case of grouped normal data, the SHEPPARD correction of the crude moments is shown to have a very high efficiency, as compared to recent attempts to improve such fits by making $\chi^2$ a minimum; the reason being that $\chi^2$ is an expression only approximate to a true value derivable from likelihood. As a final illustration of

the scope of the new process, the theory of the estimation of micro-organisms by the dilution method is investigated.

Finally it is a pleasure to thank Miss W. A. Mackenzie, for her valuable assistance in the preparation of the diagrams.

## References.

(1) K. Pearson (1920). "The Fundamental Problem of Practical Statistics," 'Biom.,' xiii., pp. 1–16.

(2) F. Y. Edgeworth (1921). "Molecular Statistics," 'J.R.S.S.,' lxxxiv., p. 83.

(3) G. U. Yule (1912). "On the Methods of Measuring Association between two Attributes," 'J.R.S.S.,' lxxv., p. 587.

(4) Student (1908). "The Probable Error of a Mean," 'Biom.," vi., p. 1.

(5) R. A. Fisher (1915). "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," 'Biom.," x., 507.

(6) R. A. Fisher (1921). "On the 'Probable Error' of a Coefficient of Correlation deduced from a Small Sample," 'Metron.," i., pt. iv., p. 82.

(7) R. A. Fisher (1920). "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and by the Mean Square Error," 'Monthly Notices of R.A.S., lxxx., 758.

(8) E. Pairman and K. Pearson (1919). "On Corrections for the Moment Coefficients of Limited Range Frequency Distributions when there are finite or infinite Ordinates and any Slopes at the Terminals of the Range," 'Biom.,' xii., p. 231.

(9) R. A. Fisher (1912). "On an Absolute Criterion for Fitting Frequency Curves," 'Messenger of Mathematics,' xli., p. 155.

(10) Bayes (1763). "An Essay towards Solving a Problem in the Doctrine of Chances," 'Phil. Trans.,' liii., p. 370.

(11) K. Pearson (1919). "Tracts for Computers. No. 1: Tables of the Digamma and Trigamma Functions," By E. Pairman, Camb. Univ. Press.

(12) K. Smith (1916). "On the 'best' Values of the Constants in Frequency Distributions," 'Biom., xi., p. 262.

(13) K. Pearson (1914). "Tables for Statisticians and Biometricians," Camb. Univ. Press.

(14) G. Vega (1764). "Thesaurus Logarithmorum Completus," p. 643.

(15) K. Pearson (1900). "On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling," 'Phil. Mag.,' l., p. 157.

(16) K. Pearson and L. N. G. Filon (1898). "Mathematical Contributions to the Theory of Evolution. IV. On the Probable Errors of Frequency Constants, and on the influence of Random Selection on Variation and Correlation," 'Phil. Trans.," cxci., p. 229.

(17) R. A. Fisher (1922). "The Interpretation of $\chi^2$ from Contingency Tables, and the Calculation of P," 'J.R.S.S.,' lxxxv., pp. 87–94.

(18) K. Pearson (1915). "On the General Theory of Multiple Contingency, with special reference to Partial Contingency," 'Biom.," xi., p. 145.

(19) K. Pearson (1903). "On the Probable Errors of Frequency Constants," 'Biom.,' ii., p. 273, Editorial.

(20) W. F. Sheppard (1898). "On the Application of the Theory of Error to Cases of Normal Distribution and Correlations," 'Phil. Trans.,' A., cxcii., p. 101.

(21) J. M. Keynes (1921). "A Treatise on Probability," Macmillan & Co., London

# VII.—*A Quantitative Investigation of the Bacterial and Protozoan Population of the Soil, with an Account of the Protozoan Fauna.*

*By* D. Ward Cutler, *M.A.*, Lettice M. Crump, *M.Sc.*, and H. Sandon, *B.A.*

*Communicated by* E. J. Russell, *F.R.S.*

(Received April 6, 1922—Read June 15, 1922.)

(From the Protozoology Department, Rothamsted Experimental Station.)

Contents.

| | PAGE |
|---|---|
| Introduction | 317 |
| Species of Protozoa | 318 |
| Influence of External Conditions on Protozoa and Bacteria | 320 |
| (1) Soil Moisture | 321 |
| (2) Rainfall | 324 |
| (3) Temperature | 325 |
| (4) Farm operations | 326 |
| (5) Chemical Changes in Soil | 327 |
| (6) Physiological Influences—Auto-intoxication | 327 |
| Discussion | 327 |
| Relation between Bacteria and Protozoa | 331 |
| Periodicity | 335 |
| Summary | 337 |
| Appendix I.—Experimental Methods | 338 |
| „ II.—Table of Results | 342 |
| Bibliography | 348 |

## Introduction.

Since the first half of the nineteenth century, it has been known that certain species of protozoa will grow in a suitable medium inoculated with soil. In spite of this, however, some later workers have asserted that such organisms did not occur in normal soils, while others, though recognising their presence, denied the existence of active forms.

To Martin and Lewin (28) must be ascribed the distinction of first showing that amœbæ and flagellates are present in field soil in a trophic condition; and owing to the labours of numerous investigators, it is now known that most soils contain these organisms; but unfortunately there is little or no knowledge as to their behaviour in natural conditions.

Earlier work (8) showed, however, that an investigation of the micro-population of field soil would yield interesting and important results, especially as it was evident

 [Published November 28, 1922.]

that certain views held by soil biologists required modification; thus it has always been thought that the bacterial numbers in the soil changed very slowly, and investigators have not hesitated to take counts at monthly or even longer intervals. This assumption is incorrect, since the numbers of both protozoa and bacteria vary enormously from day to day.

These large changes appeared so important, both from the point of view of general biology and that of soil economy, that the present investigation was undertaken with the purpose of finding in what way the changes in any one group or species are related to those in another; also what influence external conditions, such as temperature, moisture and rainfall, have upon the soil micro-organisms.

The method adopted was to make systematic counts of the numbers of bacteria and of six species of protozoa in soil samples taken daily direct from the field, and, by statistical methods, to correlate these counts one with another and with data for external conditions.

Observations at shorter periods than 24 hours could not be made owing to the labour involved; but it was found possible to organise a team of workers so that soil samples were taken at the same hour (10 o'clock) for 365 consecutive days (July 5, 1920–July 4, 1921).* In this way a large body of data was obtained, which is presented in the following pages. It is recognised that with so much material there are probable interesting deductions to be made which have been overlooked by us. In an appendix, therefore, the salient figures are given to allow those interested to apply further statistical tests.

## Species of Protozoa.

Of the species of protozoa recorded, the following six (two amœbæ and four flagellates) occurred constantly in sufficient numbers to admit the application of statistical methods to the results:—

*Dimastigamœba gruberi* (Schardinger).—This species has been described in detail by Wilson (36) under the name of *Nagleria gruberi*. It is also identical with the *Amœba punctata* of Dangeard (10) and *Vahlkampfia soli* of Martin and Lewin (27).

*Species α.*—A very small limax amœba, which has, up to the present, not been identified with any species previously described. Even when fully extended it rarely exceeds 10 $\mu$ in length. Its movements are slow, and it is never observed to travel far in any one direction. More frequently short, knobby pseudopodia are formed first at one point and then at another, and withdrawn again without any progression having been effected. In the living animal the protoplasm is refractile, and does not appear to be vacuolated or to show any structure, while the nucleus is not visible and no contractile vacuole has been observed. Staining reveals the

* Our thanks are due to Dr. M. Laurie and to Mr. P. H. H. Gray for their valuable assistance in the examination of the culture plates.

nucleus as composed of a karyosome and a very thin nuclear membrane, with no peripheral chromatin. Nuclear division is promitotic, resembling that described for *A. froschi* (29), and therefore corresponding to the second limax type in GLÄSER'S (16) classification.

The cyst is large in proportion to the size of the active form, the diameter being about 8 $\mu$. The outer wall is thick and has numerous pores, with the result that the cyst contents have a pentagonal or polygonal appearance in optical section. A very thin inner wall is closely applied to the cyst contents, which contain a large refractile granule similar to that occurring in many flagellate cysts. A full account of this species will be published later.

*Heteromita* sp.—The differentiation between the various species of *Heteromita* is in many cases not easy, and owing to the large number of cultures that had to be examined each day, it has not been attempted. Much the commonest form is one agreeing closely with the description of *Bodo repens* (KLEBS, 22), but the published descriptions do not make the distinctions between this species and *Bodo globosus* (PUSCHKAREW, 30) at all clear, and possibly both species have been present.

*Heteromita lens* (*Bodo lens*, MÜLLER, *Heteromastix lens*, WOODCOCK, 37) occurred in the cultures, and was also recorded under the name *Heteromita*, but it was not sufficiently numerous to affect the results.

*Species* $\gamma$.—This is an hitherto undescribed species. It is the smallest of the soil flagellates, varying from 3 $\mu$ to 6 $\mu$ in length and from 2 $\mu$ to 3 $\mu$ in greatest breadth. The body is slightly curved and a single flagellum arises on the concave side just behind the anterior end. Its movements are very like those of *Heteromita*, since the single flagellum adheres to the glass slide and is trailed quite passively, while the hind end of the body vibrates rapidly from side to side. The nucleus is near the anterior end and no contractile vacuole has been found. A further account of this species will be published separately.

*Cercomonas* sp.—Both *C. longicauda* and *C. crassicauda* were apparently present, but no attempt was made to distinguish between them.

*Oicomonas termo* (EHRENB.).—This is probably identical with *Monas termo*, found by MARTIN in sick soils. MARTIN (25), however, was unable to find a contractile vacuole, whereas in the present species a very distinct contractile vacuole is often present at the base of the flagellum.

## SPECIES NOT COUNTED DAILY.

The dilutions employed in counting were chosen to give the maximum accuracy for the most numerous species, and although several other protozoa were recorded regularly in small numbers, the results obtained are not sufficiently accurate to justify detailed discussion. These less numerous species are the following :—

*Amœba* sp.—A binucleate amœba which occurred almost every day in numbers up

to fifty per gramme. In appearance it is identical with *A. diploidea* (Hartmann and Nägler, 18), but its life-history has not yet been completely worked out.

*Nuclearia simplex* (Cienkowsky).—This species is not included by Cash (2) in his list of British fresh-water Rhizopoda. It is distinguished from *N. delicatula* by its smaller size, the diameter of both active and cystic forms being only between 15 $\mu$ and 25 $\mu$. It was found every day, and at times its numbers rose to as many as 3,000 per gramme.

*Chlamydophrys stercorea* (Cienk.).—A much less numerous form than the preceding.

*Monas* sp.—This species occurred constantly in small numbers.

*Proleptomonas fæcicola*.—Described by Woodcock (37) as a coprozoic species. It occurred rather less regularly than the preceding forms, but occasionally in considerable numbers.

*Colpoda cucullus* (O. F. M.).—Constantly present, but rarely exceeding fifty per gramme.

Other forms recorded occasionally in the course of the year were :—

*Rhizopoda.*

*Amœba glebæ* (Dobell, 14).

*Amœba limicola* (Rhumbler, 29a).—Rarely found.

*Hartmannella hyalina* (Dangeard, 9).

*Amœba* sp.—This is probably *Arachnula impatiens* (Cienk., 13). Only living forms were examined and these resembled Dobell's description very closely in their habits and in the absence of any apparent nucleus.

*Difflugia* sp.—This would probably have been found much more frequently had it been practicable to incubate the plates for a longer period.

*Flagellata.*

*Copromonas subtilis* (Dobell, 12), *Spiromonas angusta* (Kent), *Tetramitus rostratus* (Perty), *Tetramitus spiralis* (Goodey, 17a), *Helkesimastix fæcicola* (Woodcock), *Phyllomitus amylophagus* (Klebs), *Ophidomonas* sp., *Bodo edax* (Klebs) has been obtained from the same plot of land, but was not observed at all during the course of the count.

*Ciliata.*

*Colpoda steinii*, a much less abundant species than *C. cucullus*, *Colpidium colpoda*, *Pleurotricha* sp., *Gastrostyla* sp., *Balantiophorus* sp., *Enchelys* sp.

## Influence of External Conditions on Protozoa and Bacteria.

The numbers of all the organisms counted show large fluctuations of two kinds, viz., daily and seasonal. The size of the changes that take place within so short a time as 24 hours is perhaps the most surprising fact that the experiment has revealed. Thus the three consecutive samples 246, 247 and 248 give 58·0, 14·25

and 26·25 million bacteria respectively. The changes exhibited by any of the species of protozoa are at times even larger, and affect the active and encysted stages apparently quite indifferently. It is, in fact, not uncommon for the numbers of any species to vary from a few hundred or a few thousand to 400,000 or more within 24 hours, and the percentage of these in the active condition may jump from 0 to nearly 100 in a similar interval (*e.g. Dimastigamœba*, samples 250, 251; *Cercomonas*, samples 221, 222; *Oicomonas*, samples 245, 246; etc.).

In spite of these apparently quite erratic daily changes, when the numbers are averaged over a considerable period (*e.g.* fourteen days), marked seasonal changes became apparent. The whole soil population rises to a maximum at the end of November, and falls again to a minimum at the end of February or early in March. The bacteria then rise to a further very high figure at the end of June. The protozoa are found to fall into two well-marked groups in respect of these seasonal changes. The curves for *Cercomonas* and *Dimastigamœba* (fig. 1) run practically parallel to one another, except for a short period in September, 1920, and another in June, 1921. The curve for *Heteromita* also shows a considerable similarity to these, but, unfortunately, no great reliance can be placed on the results obtained for this organism, owing to the frequency with which its numbers exceeded the maximum that could be counted with the dilutions used. The curves for sp. $\alpha$, sp. $\gamma$ and *Oicomonas* show similar parallelism to one another (fig. 2). Thus, from the end of December till the final sample, the parallelism between the curves for sp. $\alpha$ and *Oicomonas* only breaks down at a single point, while the similarity between the curves for sp. $\alpha$ and sp. $\gamma$ during the final four months is even more striking.

The various physical, chemical and biological changes which might be expected to account for these daily and seasonal changes in numbers will be considered below. In attempting to establish a relationship either between the numbers of the different organisms or between any one of them and any of the external factors, the method adopted is to arrange all the data in the form of a contingency table (Table II, p. 328). Thus, in comparing any two variants, an increase in either is denoted by a + sign and a decrease by a − sign. ++ will then indicate that both the variants are increasing together, −− that they are decreasing together, and +− or −+ that they are changing in opposite directions. If there is no relationship between the two, the total number of like signs (++ added to −−) throughout the year should be equal to the total number of unlike signs (+− added to −+). Any departure from equality will indicate a correlation between the two.

### (1) *Influence of Soil Moisture.*

It has sometimes been claimed that the numbers of soil organisms are very closely linked up with the soil moisture. Thus Engberding (14*a*) states: "Im gewachsenen Feldboden steigt und fällt die Bakterienzahl in der wärmeren Jahreszeit mit seinem Wassergehalte," and Francé (14*b*) has recently extended this dictum to the whole of

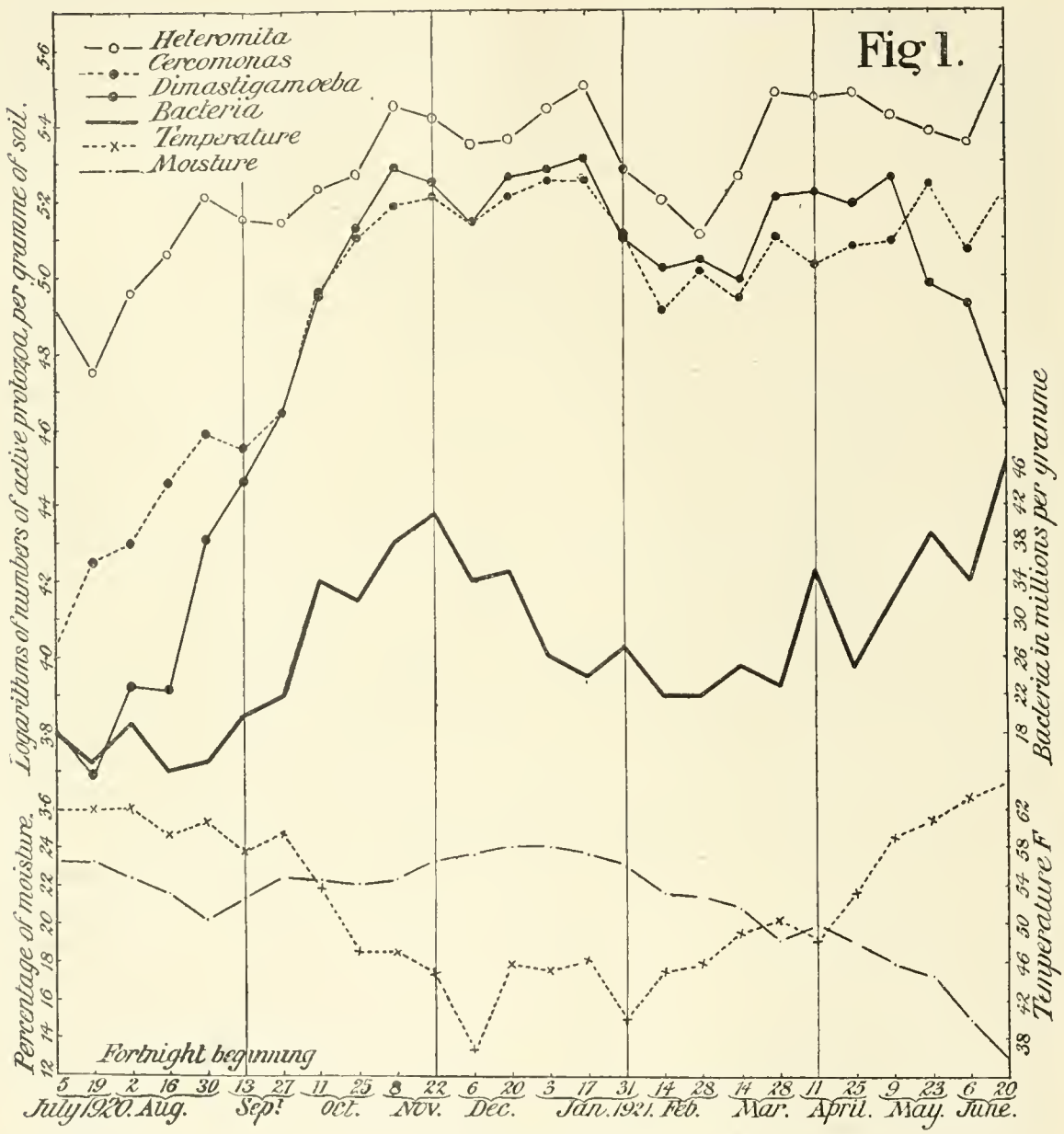


Fig 1.
Heteromita
Cercomonas
Dimastigamoeba
Bacteria
Temperature
Moisture
Logarithms of numbers of active protozoa per gramme of soil.
Bacteria in millions per gramme
Percentage of moisture.
Temperature F
Fortnight beginning
5 19 2 16 30 13 27 11 25 8 22 6 20 3 17 31 14 28 14 28 11 25 9 23 6 20
July 1920. Aug. Sept. Oct. Nov. Dec. Jan. 1921. Feb. Mar. April. May. June.

Fortnightly averages of total numbers of *Heteromita*, *Cercomonas* and *Dimastigamœba*, and of bacteria, moisture and temperature.

the soil population. The present short-interval counts, however, lend no support to this assumption. The changes in moisture-content of the soil from one day to the next are invariably small, and frequently there is no appreciable change for several successive days, while the numbers of all the organisms may be fluctuating enormously (*e.g.* samples 222, 223, 224, 225). It is impossible, therefore, to explain the daily changes in the soil population by changes in moisture, and even the seasonal moisture changes have no clear influence. There is a very striking difference in the figures obtained for nearly all the organisms at the beginning

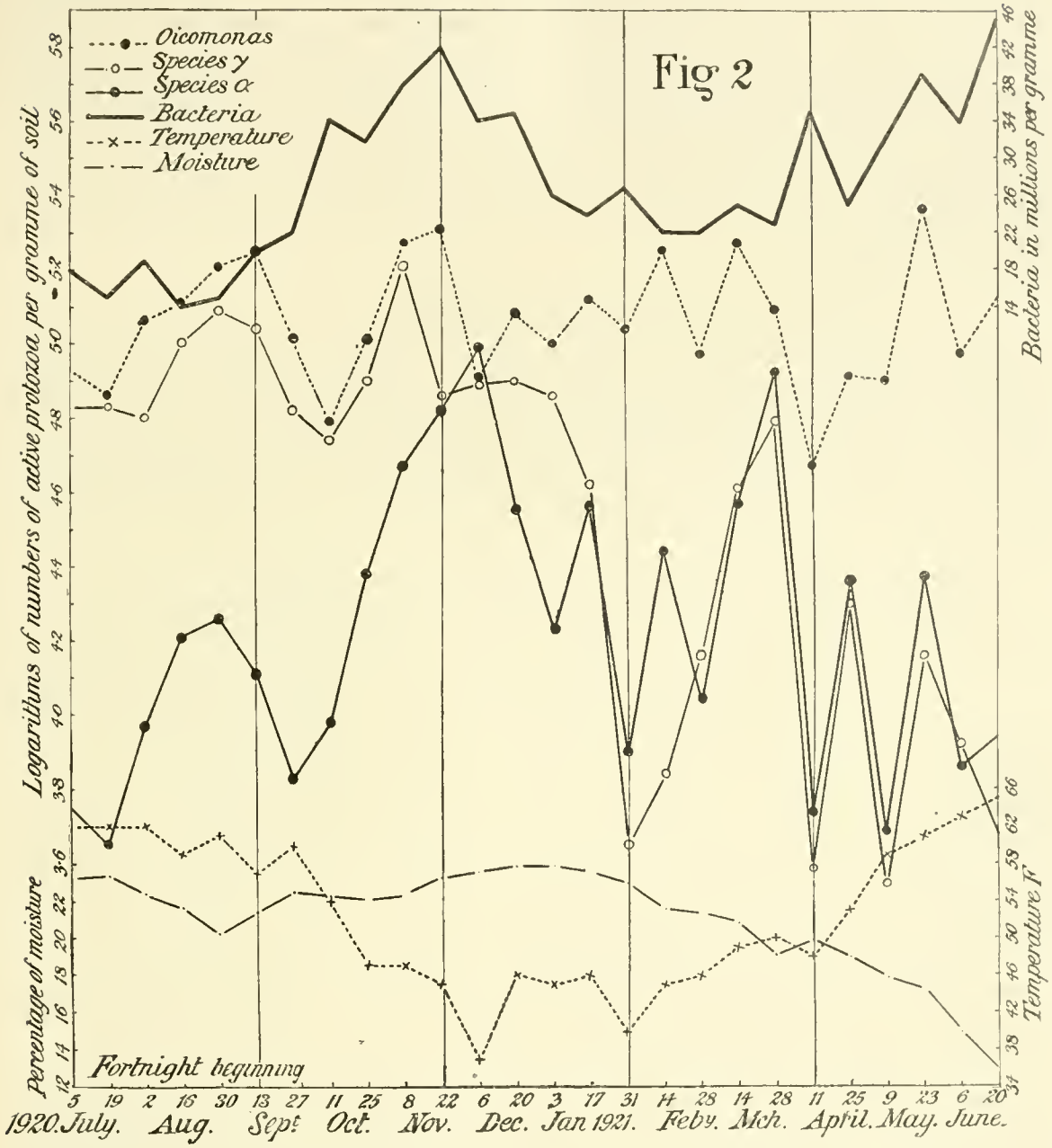


Fig 2
Oicomonas
Species γ
Species α
Bacteria
Temperature
Moisture
Logarithms of numbers of active protozoa per gramme of soil
Bacteria in millions per gramme
Percentage of moisture
Temperature F°
Fortnight beginning
5 19 2 16 30 13 27 11 25 8 22 6 20 3 17 31 14 28 14 28 11 25 9 23 6 20
1920. July. Aug. Sept. Oct. Nov. Dec. Jan 1921. Feby. Mch. April. May. June.

Fortnightly averages of total numbers of *Oicomonas*, sp. γ and sp. α, and of bacteria, moisture and temperature.

of July, 1920, from those for the end of June, 1921, and it is tempting to assume that these differences are in some way connected with the fact that the summer of 1920 was very wet, whereas 1921 was a year of unprecedented drought. Contrary to the assumptions quoted above, it is the drought which is in most cases associated with the high numbers. When the curves for average moisture and average numbers are compared in greater detail, however, it becomes impossible to draw any generalisations. Thus with the bacteria, during July and August, when the

soil was drying, the numbers were fairly constant; they rose during the following months, when the moisture was increasing, but fell considerably during December, while the moisture was still increasing. Early in April the numbers rose sharply, while the moisture-content was also rising, but during May and June, when the greatest rise in numbers occurred, the soil was drying rapidly. If the curves for any of the protozoa are considered in the same way, their independence of the moisture changes is apparent.

### (2) *Influence of Rainfall.*

It has been suggested (34) that a heavy shower of rain, by increasing the aëration of the soil solution, may have a beneficial effect on the micro-organisms living in it. Obviously, rainfall cannot be responsible for all the changes in numbers that were found, for they occurred even when no rain fell. On thirty occasions during the year the rainfall exceeded 0·2 inches, and the changes in numbers following these heavy rains are recorded in Table I.

TABLE I.—Influence of Rainfall of 0·2 inches or more on the Soil Population.

| | Bacteria. | *Dimastigamœba.* | *Sp. α.* | *Heteromita.* | *Sp. γ.* | *Cercomonas.* | *Oicomonas.* | Total Protozoa. |
|---|---|---|---|---|---|---|---|---|
| Numbers rise . | 19 | 16 | 9 | 14 | 7 | 15 | 14 | 75 |
| „ fall . | 10 | 11 | 18 | 11 | 19 | 13 | 15 | 87 |
| No change . . | 1 | 3 | 3 | 5 | 4 | 2 | 1 | 18 |
| Ratio $\frac{\text{active}}{\text{cysts}}$ :— | | | | | | | | |
| „ rises . | — | 12 | 9 | 12 | 16 | 16 | 10 | 75 |
| „ falls . | — | 12 | 18 | 11 | 12 | 11 | 18 | 82 |
| No change . . | — | 2* | 3 | 0* | 2 | 1* | 1* | 9 |

* These figures are incomplete, as it is impossible to calculate the ratio $\frac{\text{active}}{\text{cysts}}$ when the species in question was found in every plate of the series (see Appendix I, "Experimental Methods," p. 340).

These figures do not indicate any significant connection between rainfall and the daily changes in number, and this conclusion is borne out by a close examination of the records. Thus, on April 13, when the first heavy shower for over two months occurred, the bacterial numbers fell by 8·75 millions, whereas on May 8, after three weeks of dry weather, a heavy shower was accompanied by a rise of 11·1 millions. It is also noteworthy that the greater part of the autumnal rise in numbers occurred at a time when the rainfall was low, while the numbers found at the end of November, and again at the end of December and early in January (figs. 1, 2), do not seem to be influenced by the heavy rains which occurred during these periods. Conversely large changes in the average numbers sometimes occurred (as at the end of March) without any corresponding change in the rainfall.

### (3) *Influence of Soil Temperature.**

It is apparent from Table II, section 7, that rises and falls of temperature are associated quite at random with the daily changes in numbers of *Dimastigamœba* and *Cercomonas*. A similar conclusion applies to all the species of protozoa recorded. With the bacteria, the same result was established by determining the coefficient of correlation between the daily changes in the mean soil temperature and the daily change in the bacterial numbers, by the usual formula $r = \Sigma(x.y.f.)/n.\sigma_1.\sigma_2$.† The value found was $r = -0{\cdot}149 \pm 0{\cdot}01$, which is too small to indicate any connection between the two, and shows conclusively that rising temperature does not cause increased bacterial numbers.

A similar conclusion follows from an examination of the curves in figs. 1, 2. The seasonal changes in the numbers of the various organisms show no connection with the seasonal temperature changes. The big autumnal rise in numbers took place while the temperature was steadily falling, and one organism, sp. $\alpha$, actually reached its maximum at the coldest period of the year. The spring rise in numbers, on the other hand, was associated with steadily rising temperature. The unbroken rise in soil temperature between the middle of April and July does not appear to have affected the protozoal numbers, and though the average bacterial numbers rose very much during the period, the rise was broken by two large temporary falls, which show that the increase cannot be explained simply by the rising temperature.

During the fortnight, December 6 to 19, the mean soil temperature reached a minimum, owing to the occurrence of several days' continuous frost. During this period, the bacteria and all the protozoa, except sp. $\alpha$, showed a fall in numbers. A similar spell of frost occurred in February, and this time the numbers of all the protozoa were reduced, but the bacteria showed a slight rise. These observations are not very striking in themselves, but are of some interest in view of the conclusion arrived at by Russell and Appleyard (34), that the amount of $CO_2$ in the soil air is independent of temperature, except when the latter falls below 5° C., and then it is reduced. It is probable, therefore, that during a really severe winter the number of soil organisms may be considerably depressed, but owing to the mildness of the winter in which the present counts were made, the results obtained are not conclusive on this point. Even while the surface of the soil was

* The mean soil temperature has been reckoned by the method recommended by Keen and Russell (21), in which the average is found of the maximum air temperature and of the minimum recorded by the 12-inch soil thermometer during the 24 hours. The readings used were those taken daily at 10 A.M. at the Rothamsted meteorological station, situated some 200 yards from the plot; tests showed that the temperatures recorded at the two places were identical.

† $r$ = coefficient of correlation; $n$ = total number of cases; $f$ = number of cases in class; $x$ = deviation of class from mean temperature change; $y$ = deviation of class from mean change in bacterial numbers; $\sigma_1$ = standard deviation of mean temperature changes; $\sigma_2$ = standard deviation of changes in bacterial numbers.

frozen, however, the numbers of all the organisms showed large apparently random daily fluctuations just as at other times. It was only when averages were taken that the depressions became apparent.

(4) *Influence of Farm Operations.*

An increase in the numbers of bacteria in a field soil during the autumn has been observed by numerous workers in the past, and is commonly attributed either to the

aërating effect of autumn ploughing, or to the nutritive effect of manuring. Neither of these explanations apply in the present case, for none of the farm operations (see Appendix I) were followed by any abnormal changes in numbers, and no manure was applied during the years 1920 and 1921.

(5) *Influence of Chemical Changes in the Soil.*

In preparing the soil sample each day for counting protozoal cysts, 10 c.c. of the acid was always withdrawn just before the dilutions were made, and was titrated with alkali in order to ascertain that its concentration, after neutralising the carbonate of the soil, was still sufficient to kill all the active protozoa. Slight differences were found from day to day in the amount of alkali required for the titration, and these must correspond to differences in the chemical conditions of the soil samples, but no connection could be found between them and the changes in the number of any of the organisms.

(6) *Physiological Influences—Autointoxication.*

A population of 500,000 individuals of any species of protozoa per gramme of soil represents about 2,500,000 per c.c. of soil solution, for 20 per cent. may be taken as a typical value of the water content of Rothamsted soil. If, as is not impossible, the organisms are not distributed uniformly throughout this medium, but are aggregated at some particular points among the soil particles, their density may be sufficiently high for the accumulation of their waste products, in the absence of constant renewal of the solution, to exert a depressing effect on the numbers as commonly occurs in artificial cultures. During the year, it was sometimes noticed that exceptionally high numbers of any organism on any day were followed by low numbers the next day, as if autointoxication had taken place. But an examination of the complete figures for the whole year does not substantiate this hypothesis.

## Discussion.

From the foregoing sections it appears that, contrary to all expectations, no connection can be traced between climatic conditions or the other external factors considered and either the daily or the seasonal changes in the numbers of any of the soil organisms investigated. It is, however, difficult to resist the conclusion that some such connection must exist. The similarity between the seasonal curves for all the different organisms suggests that they are reacting together at any rate to the larger climatic changes. There is, moreover, small but significant correlation between the daily changes of some of the protozoa. Table II, section 2, shows that there is a definite excess of like signs when the daily changes in any two flagellates are taken together, and also when *Dimastigamœba* is taken with either *Cercomonas* or *Heteromita*.

The similarity that was observed between the seasonal changes in *Dimastigamœba*,

TABLE II.—Contingency Tables for various Soil Organisms.

(For explanation, see p. 321.)

| | + + | − − | + − | − + | Total No. of | | Probable error. |
|---|---|---|---|---|---|---|---|
| | | | | | Like signs. | Unlike signs. | |
| 1. Between active amœbæ and baeteria— | | | | | | | |
| Inverse relationship— | | | | | | | |
| *Dimastigamœba* and bacteria | 47 | 54 | 124 | 114 | 101 | 238 | ± 9 |
| *Sp. α* and bacteria | 79 | 87½ | 95 | 98½ | 166½ | 193½ | ± 9½ |
| *Dimastigamœba* + *Sp. α* and bacteria | 39½ | 44 | 128½ | 127 | 83½ | 255½ | ± 9 |
| 2. Between numbers of active individuals of different species of protozoa on the same day— | | | | | | | |
| Direct relationship indicated— | | | | | | | |
| *Oicomonas* and *Cercomonas* | 56 | 46½ | 40 | 32½ | 102½ | 72½ | ± 6½ |
| „ „ *Heteromita* | 43½ | 39 | 27½ | 19 | 82½ | 46½ | ± 5½ |
| „ „ *Sp. γ* | 82 | 78 | 66 | 43 | 160 | 109 | ± 8 |
| *Sp. γ* and *Heteromita* | 78½ | 77½ | 47½ | 45½ | 156 | 93 | ± 8 |
| „ „ *Cercomonas* | 92 | 98 | 66 | 75 | 190 | 141 | ± 9 |
| *Heteromita* and *Cercomonas* | 60 | 73 | 42 | 54 | 133 | 96 | ± 7½ |
| *Dimastigamœba* and *Cercomonas* | 104¼ | 109¾ | 76¼ | 69¾ | 214 | 146 | ± 9½ |
| „ „ *Heteromita* | 87 | 77½ | 61 | 64½ | 164½ | 125½ | ± 8½ |
| No relationship apparent— | | | | | | | |
| *Dimastigamœba* and *Sp. α* | 86½ | 91½ | 84½ | 78½ | 178 | 163 | ± 9 |
| „ „ *Sp. γ* | 92 | 91 | 91½ | 86½ | 183 | 178 | ± 9½ |
| *Sp. γ* and *Sp. α* | 89½ | 91½ | 92½ | 86½ | 181 | 179 | ± 9½ |
| *Sp. α* and *Oicomonas* | 52½ | 37½ | 45½ | 43½ | 90 | 89 | ± 6½ |
| 3. Between numbers of cysts of different species of protozoa on the same day— | | | | | | | |
| Direct relationship indicated— | | | | | | | |
| *Oicomonas* and *Cercomonas* | 102 | 114½ | 61 | 72½ | 216½ | 133½ | ± 9½ |
| „ „ *Heteromita* | 74 | 93 | 41 | 57 | 167 | 98 | ± 8 |
| „ „ *Sp. γ* | 98½ | 123½ | 67½ | 71½ | 222 | 139 | ± 9½ |
| *Sp. γ* and *Heteromita* | 56 | 63 | 45 | 48 | 119 | 93 | ± 7 |
| „ „ *Cercomonas* | 89 | 100½ | 85½ | 77 | 189½ | 162½ | ± 9½ |
| *Heteromita* and *Cercomonas* | 80 | 80 | 45½ | 53½ | 160 | 99 | ± 8 |
| *Dimastigamœba* and *Sp. α* | 98¼ | 102¼ | 75¾ | 82¾ | 200½ | 158½ | ± 9½ |
| No relationship apparent—*Dimastigamœba* and *Sp. γ* | 83¾ | 95¼ | 87½ | 95½ | 179 | 183 | ± 9½ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4. Between total numbers (active and cystic) of different species of protozoa on the same day— | | | | | | | |
| Direct relationship indicated— | | | | | | | |
| *Dimastigamœba* and *Sp. α* | 96 | 95½ | 77½ | 79 | 191½ | 156½ | ± 9½ |
| „ „ *Cercomonas* | 110¾ | 110¼ | 71¾ | 69¼ | 221 | 141 | ± 9½ |
| No relationship apparent—*Dimastigamœba* and *Sp. γ* | 93½ | 96½ | 89½ | 85½ | 190 | 175 | ± 9½ |
| 5. Between numbers of a single species of protozoa on successive days— | | | | | | | |
| Inverse relationship— | | | | | | | |
| *Oicomonas:* | | | | | | | |
| Actives of 1st day and actives of 2nd day | 20¾ | 33¾ | 151¾ | 152¾ | 54½ | 304½ | ± 9 |
| „ „ total „ | 43¾ | 57¼ | 129¾ | 128¼ | 101 | 258 | |
| Total „ „ „ | 41½ | 56 | 130½ | 130 | 97½ | 260½ | ± 9 |
| *Dimastigamœba:* | | | | | | | |
| Active of 1st day and active of 2nd day, etc. | 43½ | 48½ | 130½ | 129½ | 92 | 260 | ± 9 |
| Direct relationship— | | | | | | | |
| *Oicomonas:* | | | | | | | |
| Active of 1st day and cystic of 2nd day, etc. | 88½ | 110½ | 85 | 75 | 199 | 160 | ± 9½ |
| *Heteromita:* | | | | | | | |
| Active of 1st day and cystic of 2nd day, etc. | 71½ | 58½ | 44½ | 36½ | 130 | 81 | ± 7 |
| No relationship apparent— | | | | | | | |
| *Oicomonas:* | | | | | | | |
| Cystic of 1st day and cystic of 2nd day, etc. | 46¼ | 73¼ | 119¾ | 118¾ | 119½ | 238½ | ± 9 |
| *Cercomonas:* | | | | | | | |
| Active of 1st day and active of 2nd day, etc | 49¼ | 62¼ | 120¾ | 117¾ | 111½ | 238½ | ± 9 |
| „ „ cystic „ | 86 | 91 | 85 | 93 | 177 | 178 | ± 9½ |
| *Heteromita:* | | | | | | | |
| Active of 1st day and active of 2nd day, etc. | 41¼ | 31¼ | 74¾ | 94¾ | 72½ | 149½ | ± 7 |
| *Sp. γ:* | | | | | | | |
| Active of 1st day and active of 2nd day, etc. | 54¼ | 57¾ | 123¼ | 125¾ | 112 | 249 | ± 9 |
| „ „ cystic „ | 83¼ | 93¾ | 98¼ | 85¾ | 177 | 184 | ± 9½ |
| *Dimastigamœba:* | | | | | | | |
| Active of 1st day and cystic of 2nd day, etc. | 93½ | 93 | 89 | 84½ | 186½ | 173½ | ± 9½ |
| Cystic „ „ „ | 63½ | 65½ | 120 | 115 | 129 | 235 | ± 9 |
| Total „ total „ | 63 | 61 | 118 | 115 | 124 | 233 | ± 9 |
| *Sp. α:* | | | | | | | |
| Active of 1st day and active of 2nd day | 52 | 61 | 124½ | 126½ | 113 | 251 | ± 9 |
| „ „ cystic „ | 79¼ | 89¼ | 97¾ | 94¾ | 168½ | 192½ | ± 9½ |
| 6. Between numbers of active and cystic *Oicomonas* on the same day | 65¾ | 89¾ | 107¾ | 98¾ | 155½ | 206½ | ± 9½ |
| 7. Between the total numbers (active and cystic) and temperature— | | | | | | | |
| No relationship apparent— | | | | | | | |
| *Cercomonas* and temperature | 98 | 75½ | 85½ | 104 | 173½ | 189½ | ± 9½ |
| *Dimastigamœba* and temperature | 96½ | 75½ | 85½ | 104½ | 172 | 190 | ± 9½ |

*Cercomonas* and *Heteromita* therefore extends also to some extent to the daily changes both in their cystic and active numbers. On the other hand although the seasonal curve for sp. $\alpha$ is so strikingly parallel to those for sp. $\gamma$ and *Oicomonas* the daily fluctuations of this amœba are quite independent of the other two species. The absence of a preponderance of unlike signs, however (except in the case of the amœbæ and bacteria which will be considered separately, p. 331), shows clearly that there is no antagonism between any of the different organisms. It appears, therefore, that, though all the organisms respond alike to seasonal environmental changes, their responses to the more temporary environmental changes are much less uniform. From the available data, however, it is impossible to say what is the actual connection between the changes in numbers of any of the soil organisms and the changes in their environment. In view of the magnitude of the changes in numbers that occur from one day to the next, it is obviously desirable to obtain some knowledge of the changes that take place within the course of a single day and night; but this is at present impracticable, and probably the only way of advance at present is through an extensive series of laboratory experiments under rigidly controlled conditions.

It is interesting to note that the seasonal variations in the numbers of the soil organisms is very similar to those which have frequently been recorded in the case of many aquatic organisms. Miss Delf (11), for instance, found that in the ponds at Hampstead the algæ are most numerous in the early spring (February to the beginning of May) and again in October and November, some species disappearing entirely during the intervals. Somewhat similar seasonal changes have been recorded in a pond by Fritsch and Rich (15), in the British lakes by West and West (35), in the Illinois River by Kofoid (22A), and elsewhere.

In the sea also the planktonic diatoms have a great pulse of activity between March and June, and the dinoflagellates attain their maximum about a month later. Both groups have a smaller second maximum in the late autumn (19).

Thus the maxima for the soil population coincide closely with those of the marine and fresh-water plankton, except that whereas the aquatic forms are most active in spring and only show a minor phase of activity in autumn, the reverse appears to be true in the soil. It is difficult to resist the conclusion that the annual variations are produced by similar causes in each case. If this assumption is correct, it follows that the seasonal changes in protozoal numbers in the soil are not determined solely by the number of bacteria present, in spite of the fact that bacteria are probably the only food supply of the species in question, for the aquatic algæ which show such closely similar seasonal changes are not dependent on such a food supply. This conclusion is supported by the fact that the number of all the protozoa except *Oicomonas* rose during March, whereas the corresponding increase in the bacteria was delayed till the beginning of April.

The annual fluctuations in aquatic organisms agree further with those in the soil organisms in having little if any close relation with temperature. A marked annual

periodicity was found by Apstein (1) even in a lake in Ceylon, where the total range of temperature during the year was from 25·94° C. in November to 28·28° in May. On the other hand, the fresh-water organisms are influenced very considerably by rainfall and general hydrographic conditions (see especially Apstein (1) and Kofoid (22A)), whereas no such influence could be found in the soil. Solar energy also undoubtedly influences the development of the phyto-plankton (15), and Herdman (19) regards it as being probably the chief cause of the great increase of life in the sea in the spring.

It is interesting to note that seasonal changes very similar to those found in the soil bacteria have been recorded for sewage bacteria. At the Dorking sewage works, Dr. (now Sir) A. C. Houston (31) found the highest numbers during October, November and December, while during February, March and April the numbers were consistently lower than at any other time of the year.

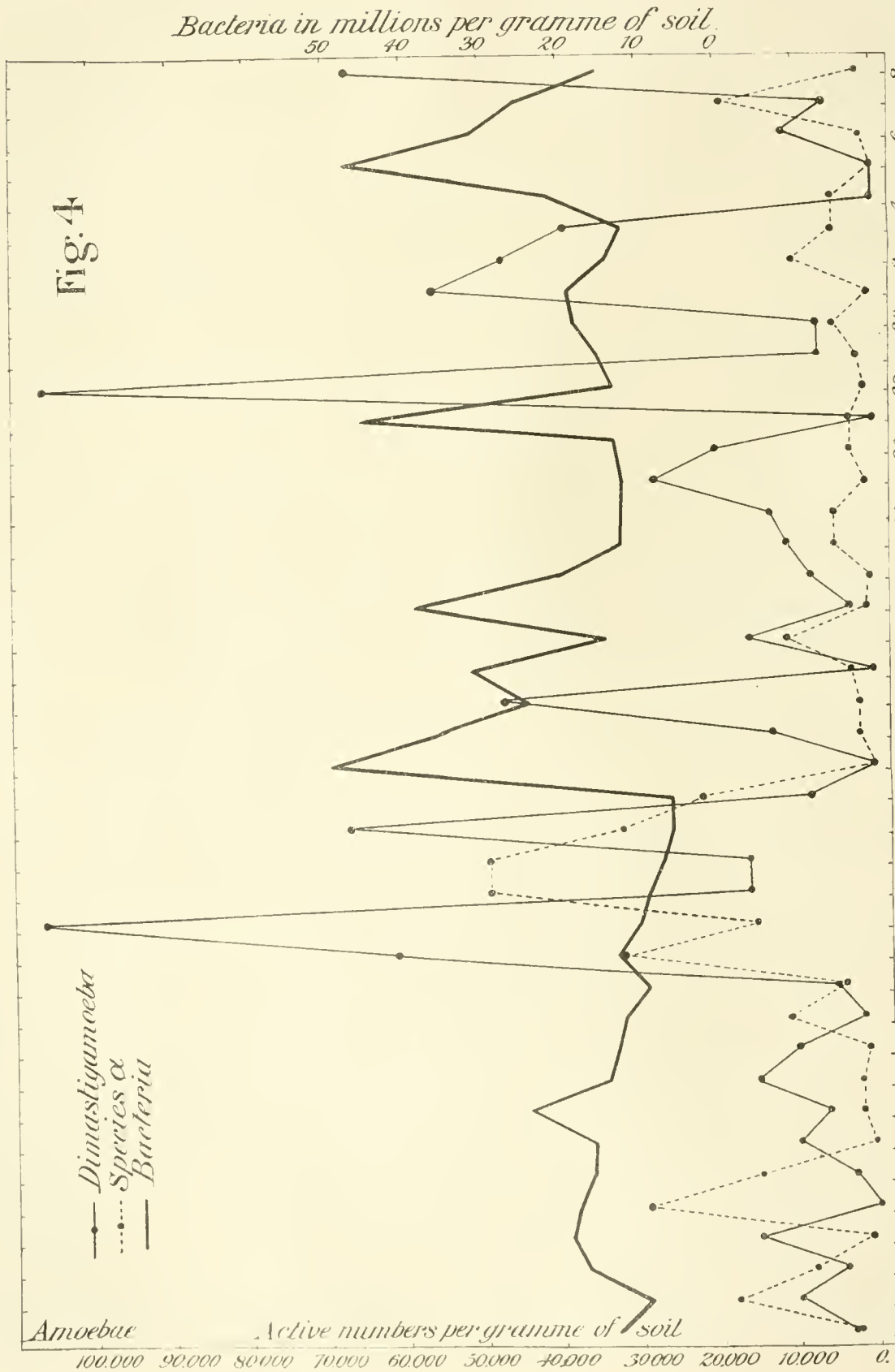
### Relation between Protozoa and Bacteria.

Comparison of the numbers of bacteria with those of the active amœbæ (*Dimastigamœba* and sp. α) shows a definite inverse relationship between them. A contingency table (Table II, sec. 1) formed from these numbers shows that they vary inversely 255·5 times and together 83·5 times, with a probable error ±9·2 (figs. 4, 5, and 6, Samples 320–327). If the numbers of active individuals of *Dimastigamœba* and sp. α are considered separately with the bacteria, the result is as follows:—

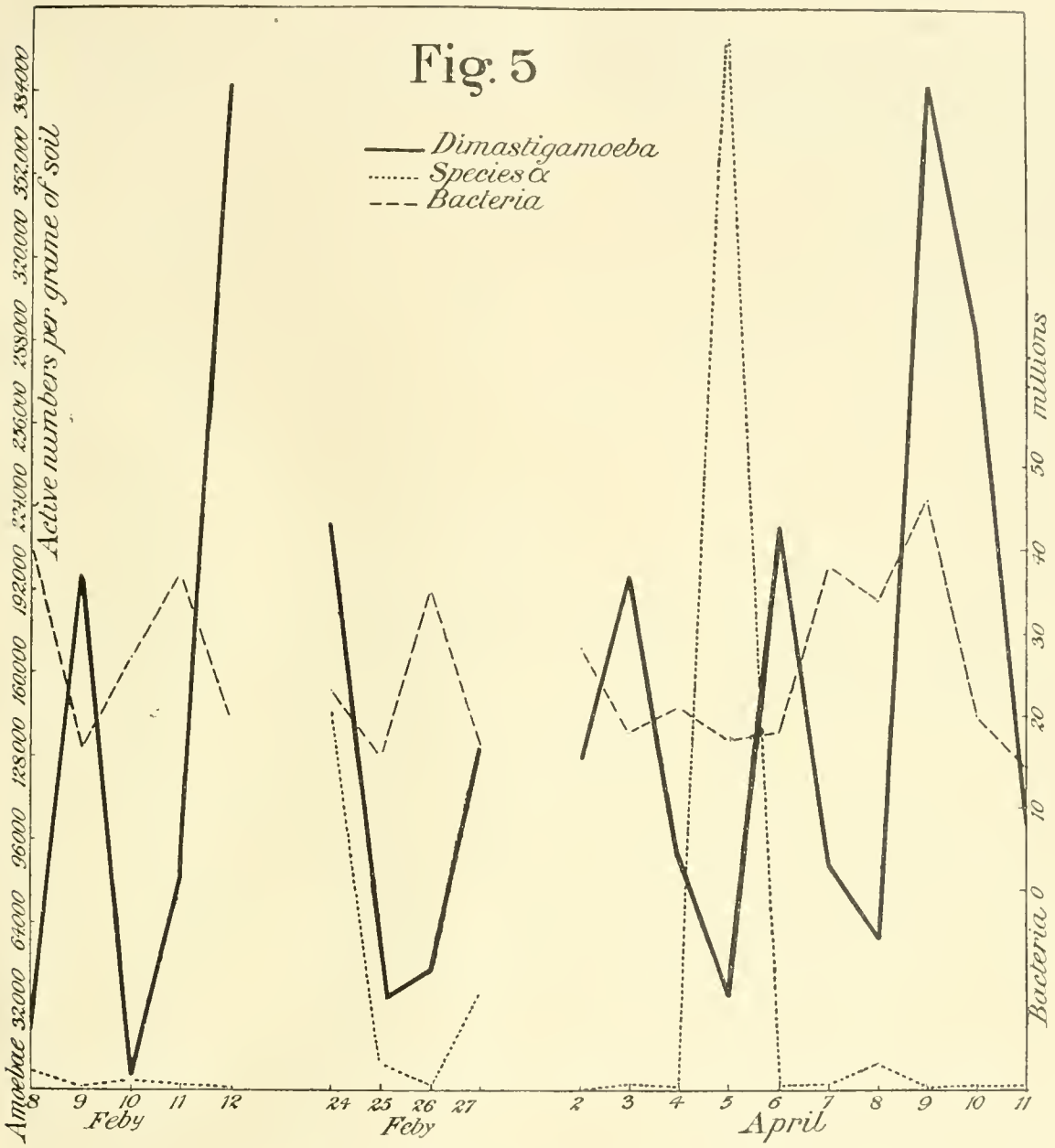
| | Like. | Unlike. |
|---|---|---|
| *Dimastigamœba* and bacteria . . . | 101 | 238 (70 per cent. unlike). |
| Sp. α and bacteria . . . . . | 166·5 | 193·5 (54 per cent. unlike). |

In both cases the difference in numbers is significant, but obviously *Dimastigamœba* is very much more effective in its action on bacteria than is sp. α; but in fifty-three cases it has either remained unchanged or varied with bacteria, while sp. α has varied inversely with them. Such cases occur at Samples 34, 41, 49, etc. (fig. 6, Samples 162, 163). An examination of the data shows that generally, where the influence of sp. α on the bacterial numbers is apparent, the numbers of *Dimastigamœba* are either changing very slightly (in twenty of the fifty-three cases they have not changed at all), or else the changes in sp. α are large (*e.g.* Samples 137, 275).

Out of the 339 days considered in the formation of these contingency tables, *Dimastigamœba* shows a controlling influence on the bacteria 70 per cent. times, and sp. α another 16 per cent. There remain then 14 per cent. or forty-eight days on which the changes in numbers in the bacteria cannot apparently be attributed to the action of the amœbæ, since they vary directly and not inversely with each other. Certain of these can be explained by reference to the amœbæ if the cysts as well as the active forms are taken into consideration. At Sample 28 the numbers of sp. α are practically unchanged, but the bacteria and *Dimastigamœba* have risen; in the

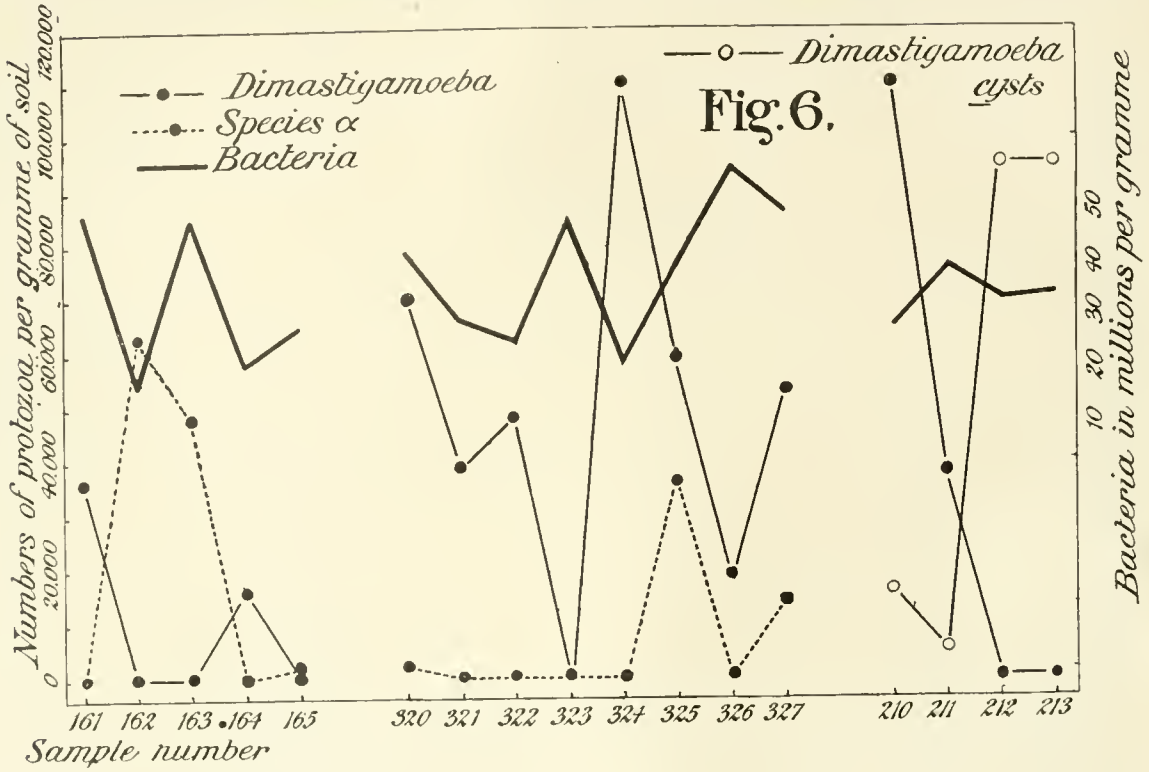

Bacteria in millions per gramme of soil
50 40 30 20 10 0
Fig. 4
Dimastigamoebæ
Species α
Bacteria
Amoebae
Active numbers per gramme of soil
100,000 90,000 80,000 70,000 60,000 50,000 40,000 30,000 20,000 10,000 0.
29 31 August 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 2 4 6 8 October

Daily numbers of active amœbæ (*Dimastigamœba* and sp. *α*) and bacteria in 1 grm. of field soil, from August 29 to October 8, 1920.




Numbers of active amœbæ (*Dimastigamœba* and sp. α) and bacteria in 1 grm. of field soil for typical periods in February and April.

latter, however, the rise is entirely due to excystment, so that the influence upon the bacteria would naturally have been slight, especially if the excystment did not occur until near the time of sampling. At Sample 35 sp. α is again unchanged, the bacteria fall, and in *Dimastigamœba* the active numbers fall slightly, though the totals rise by 8000, showing that considerable activity has gone on in the preceding twenty-four hours, terminating in cyst formation. At Sample 212 (fig. 6) there is a similar case; the bacteria and active *Dimastigamœbæ* fall, and though sp. α rises, the



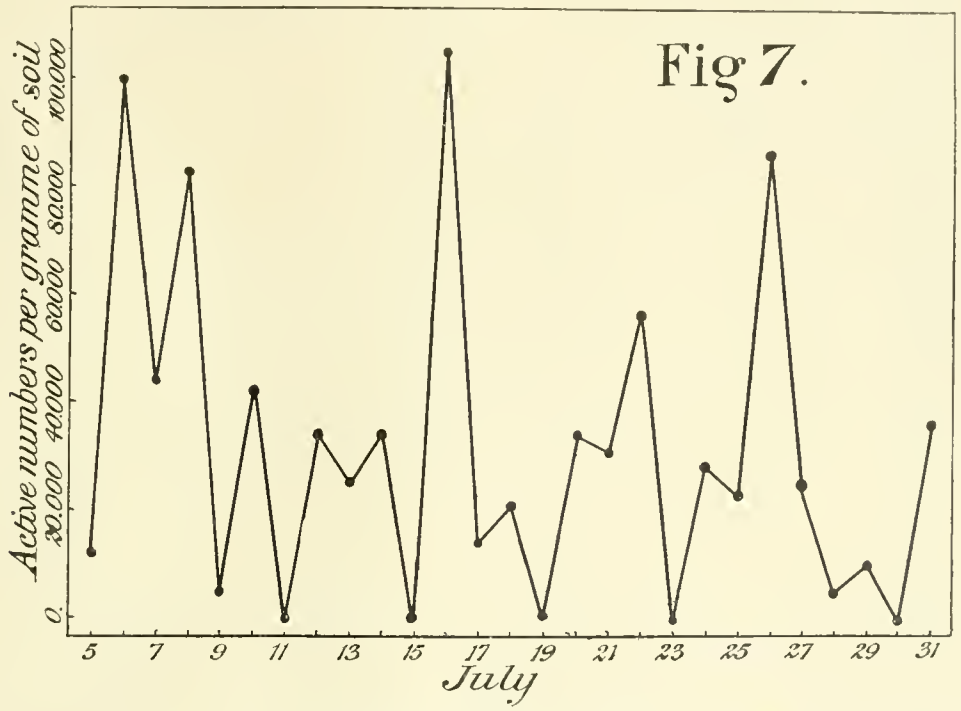

Numbers of amœbæ and bacteria per gramme of soil for short periods in December, 1920, and May and February, 1921.

increase (1000) is scarcely enough to account for the bacterial change, but there has been a large increase in the cysts of *Dimastigamœba* (97,000). At Sample 297 there is an increase in both active amœbæ and in the bacteria, but in the amœbæ the change is due to excystment, as in Sample 28. Sample 321 resembles 35, but in this case it is the cysts of sp. α which have increased by 11,000. At Samples 362 and 363 sp. α may be disregarded, as the numbers are almost stationary: the bacteria fell on both days, and on both there were no active *Dimastigamœbæ* when the samples were taken. Again, there is evidence of activity during the time between the counts, as Sample 362 shows a fall in both cysts and totals, so death is not likely to have taken place until excystment and consequent activity had occurred, while Sample 363 shows a rise in cysts and totals, again implying that there has been an intervening period of activity. There are numerous other cases where the same thing has occurred, but in these the active numbers at the time of sampling are sufficient to account for the bacterial changes.

From an inspection of the figures it is not possible to decide whether the interaction is primarily one in which the amœbæ influence the bacteria or *vice versâ*. The possibility of the bacterial changes initiating those in the amœbæ is, however, entirely ruled out by the fact that the contingency tables show no positive correlation between the changes in numbers of the active forms in the two amœbæ.

### Periodicity.

Preliminary experiments (8) had shown a remarkable periodic fluctuation in the numbers of active flagellates in the soil, high numbers on one day being succeeded by low, which were again followed by high on the third day. The extended observations here described have confirmed this phenomenon and have shown it to be characteristic of only one species of flagellates, viz., *Oicomonas termo* (Ehren.). The rhythm has been maintained, with few exceptions, for the 365 days and is shown in figs. 7 and 8 for different periods of the year.




Daily variations in the numbers of active individuals of a species of soil flagellate, *Oicomonas termo* (Ehrenb.), during July, 1920.

It can be shown that if there were no periodicity, and if the fluctuations were purely random, then the number of like signs in a contingency table made for the active numbers on successive days would be $33\frac{1}{3}$ per cent. and the number of unlike signs $66\frac{2}{3}$ per cent. of the total observations. The observed figures give a percentage of only $15{\cdot}08 \pm 1{\cdot}89$ like signs, which is highly significant (see Table II, sec. 5).

The possibility was considered that an alternation of encystation and excystation might partially account for the variation of the active forms; this, however, does not hold, for the cyst curve is not the inverse of the active one. Also, if the hypothesis were correct, the cysts should show a periodicity. To test this possibility the numbers of cysts on successive days were compared, but gave $33{\cdot}43 \pm 2{\cdot}49$ per cent. like signs, which is practically the expectation, if excystation and encystation are wholly non-periodic. The further test was made of comparing active forms of

one day with the cystic of the next. The result was 199 like, 160 unlike signs. Now, if there is no connection between the two, that is, if high active numbers do not tend to cause cyst formation, or, conversely, the expectation would be 50 per cent. like to 50 per cent. unlike signs, which is not widely different from the

Daily variations in the numbers of active individuals of a species of soil flagellate, *Oicomonas termo* (EHRENB.), during March, 1921.

numbers obtained. There is, however, a small significant excess of like signs, brought about by the tendency for a drop in active numbers, to be followed by a similar drop in the number of cysts (see Table II, sec. 5). The significance of this tendency is, however, sufficiently small to render it almost negligible. Thus the hypothesis of regular excystation and encystation will not completely account for the periodicity of the active forms. As would be expected from the above considerations, the total numbers of *Oicomonas* are periodic to a small extent. Thus the comparison of the totals for successive days gives 27·33 per cent.$\pm$2·35 of like signs, which is significant, though not to as high a degree as that of the active forms. Also, as is seen in Table II, sec. 5, the unlike signs predominate when the actives and totals for successive days are compared.

It may be deduced, therefore, that increased active numbers tend to be followed by decreased totals, that is death, conjugation, or both, while falls in the active numbers are followed by rises in the totals—reproduction.

It has been found impossible to correlate any obvious external condition with the rhythm, indeed, it is difficult to conceive of a periodic external condition such as would be necessary; and, moreover, pure cultures in artificial media maintained at constant temperature in the laboratory, exhibit a precisely similar periodicity of active numbers.

It is concluded, therefore, that the solution of the problem must be sought in the organism itself, though it is realised that environmental conditions, such as food supply, will play a part in the complexity of factors involved.

### *Other Species.*

The active numbers of *D. gruberi*, when compared for successive days, also show a tendency for the active forms to alternate from day to day. But it is not so marked as in the case of *Oicomonas*, since the unlike signs are only 76·7 per cent., as against the expected 66·6 per cent. The other species show no tendency to periodicity when tested for a two days' period, for their active numbers are found distributed in a random way. It is possible that a rhythm of a longer interval—three, four, or five days—may obtain, but this can be tested only by experiment in controlled culture media.

In no case is there evidence that cyst formation occurs regularly; in every species considered, the results of comparisons between actives and cystic, or between the cystic numbers for alternate days, are what would be expected on the supposition of random excystation and encystation, or the departure from expectation is too small to be worth comment. Thus, although excystation and encystation play a part in the variation of the active numbers, the preponderating influence is death, conjugation and reproduction, causing variation in the total numbers of the species.

## Summary.

1. The results of 365 consecutive daily counts of the numbers of bacteria and of six species of protozoa in a normal field soil are given, and the methods of counting bacteria and protozoa are described.

2. The numbers of both bacteria and protozoa rarely remain the same from one day to the next. The fluctuations are very great, but it has not been found possible to connect them with meteorological or general soil conditions.

3. Fourteen-day averages of the daily numbers demonstrate that well-marked seasonal changes in the soil population are superimposed on the daily variations in numbers. In general, both bacteria and protozoa are most numerous at the end of November and fewest during February. These changes are not directly influenced by temperature or rainfall, but show a similarity to the seasonal fluctuations recorded for many aquatic organisms.

4. There is a slight tendency for the various species of flagellates to fluctuate together from day to day, but this is not shown by the two species of amœba.

5. An inverse relationship is found between the numbers of bacteria and active amœbæ in 86 per cent. of the total observations.

6. A two-day periodicity obtains for the active numbers of one species of flagellate, *Oicomonas termo*.

## APPENDIX I.

### Experimental Methods.

#### *Sampling of Soil.*

The samples of soil were taken from the plot on Barnfield, Rothamsted, which annually receives farmyard manure at the rate of 14 tons to the acre. The portion of the plot chosen was approximately 20 yards long by 12 yards broad. No crop was grown during the period of the experiment, and no manure added in the year 1920; but, in July–August, 1920, it was horse-hoed; on February 5–7, 1921, ploughed; on March 3, cultivated; on April 15–16, 27–28, rolled; and in May, June, and July, hoed.

The samples were taken at the same time (10 a.m.) every day, the method being as follows:—

Six 9-inch borings, taken at intervals of about 6 feet in a line across the plot, and thoroughly mixed together formed Sample 1, which was taken to the laboratory in a sterile tin.

The second sample was taken on the following day in the same manner and on the same line as the first, except that the borings were made about 2 feet away. The third sample borings were made about 2 feet from those of the second sample, thus completing the first line across the plot. (See Map, p. 339.) For the next three samples a second line about 9 inches from the first was used, and so on to the end of the series. This method of sampling was adopted to obviate, as far as possible, errors due to unequal distribution of the micro-organisms over the plot, also to ensure against such an error control experiments were made. On three occasions four samples were taken simultaneously across the plot, and counted. The results for the protozoa and bacteria were in no case significantly different in the four samples. Further, for the last fifteen days a duplicate sample was taken 6 yards back, and the bacteria counted;

the results of the two counts were concordant. These experiments justify the belief that unequal distribution in the field cannot account for the daily fluctuations in number of the organisms.

Map showing method of taking soil samples, each sample being formed from six borings, and each boring being denoted by the respective numeral.

*Bacteria.*

The nutrient medium employed is a modification, prepared by Mr. H. G. Thornton and as yet unpublished, of one described by Conn. It consists of:—

| | | |
|---|---|---|
| Distilled water . . . . . | 1 | litre. |
| $K_2HPO_4$ . . . . . | 1 | grm. |
| $KNO_3$ . . . . . | 0·5 | ,, |
| $MgSO_4$ . . . . . | 0·2 | ,, |
| $CaCl_2$, KCl, NaCl . . . . . | 0·1 | ,, each. |
| Asparagine . . . . . | 0·5 | ,, |
| Agar . . . . . | 15·0 | ,, |
| $FeCl_3$ . . . . . | 0·002 | ,, |

After the substances have melted in a steamer 1 grm. of mannite is added, and the $P_H$ value adjusted to 7·3.

Since this medium contains only simple chemical substances it is not liable to the variations which often occur in successive preparations containing peptone, lemco, or gelatine. Such variations have been found to be sources of considerable error in counting.

The bacteria are counted by a dilution method, practicable direct means being yet unknown; 10 grm. of soil, which have passed through a 3-mm. sieve, are placed in 250 c.c. of sterile physiological salt solution, contained in a 750 c.c. flask, and shaken for 5 minutes, thus giving a 1/25 dilution. 1 c.c. of this is shaken for 1 minute in 100 c.c. of saline contained in a 300 c.c. flask; 1 c.c. of this second dilution is shaken for 1 minute in a 100 c.c. of saline contained in a 300 c.c. flask, thus giving a dilution of 1/250,000. Of this 1 c.c. quantities are placed in the bottom of five separate sterile Petri dishes; 5 c.c. of the nutritive medium are added, and a thorough admixture ensured by rotating the dish. After incubation at 20° C. for 7 days the total number of colonies are counted, and the numbers of bacteria in millions per gramme calculated.

Strict adherence to the details of the technique is necessary, otherwise the results are not reliable.

The data obtained have been examined statistically by Mr. R. A. Fisher, and in nearly 80 per cent. of the 365 daily counts of bacterial numbers, the distribution of variation in the numbers of colonies on each plate is strictly that which should be obtained by random sampling as calculated by the Poisson series. This series represents the mathematical consequences of an accurate technique of sampling and dilution. That the 10 per cent. of exceptions (marked with an asterisk in Table IV) are not extreme cases of systematic errors affecting the whole of the data, is shown by the sharpness with which they may be distinguished from the rest by the statistical tests, and by the conformity of the 90 per cent. to the theoretical distribution. For these latter counts the probable error would therefore be approximately 2/3 of the standard error, which depends solely on the total number of colonies counted. Thus, if the total number of colonies counted on a series of plates is 291, this figure is affected by a standard error of $\sqrt{291} = 17{\cdot}1$, or by a probable error $17{\cdot}1 \div 2/3 = 11{\cdot}4$. Since this example was obtained from a four-plate series, the result in millions per gramme is obtained by dividing by 16 giving $18{\cdot}2 \pm 0{\cdot}71$.

The probable error varies with each count made, but for the whole series (excluding the exceptions) it is between 4–6 per cent.

*Protozoa.*

The protozoa are counted by a dilution method, the essentials of which have already been described (7). In the present experiment the three ranges of dilutions employed at different times of the year (see Table IV) were 1/25–1/102,400, 1/50–1/204,800 and 1/50–1/409,600. It has been shown (6, 7) that a more complete census of the soil population is obtained if the Petri dishes, inoculated from each dilution, are incubated for twenty-eight days and examined weekly, since a few protozoa, *e.g. Chlamydophrys*, develop only slowly; but it was impracticable to continue observations for so long owing to the great labour involved. The counts, therefore, refer only to those species which reproduce sufficiently rapidly to give accurate results after fourteen days' incubation. Even with this restriction over 100 plates were examined daily, and there were always about 700 plates in the incubator room.

The numbers of organisms in the soil sample are deduced as follows:—

For a given soil population the average number of negative plates in a series (that is plates showing no protozoan development) can be calculated, consequently any observed number of negative plates will represent a certain density of population in the soil sample investigated. Such calculations have been done by Mr. R. A. Fisher who compiled Table III for us. Thus in a series of dilutions ranging from 1/25–1/102,400 five negative plates will represent an average population of 16,000 per gramme of soil. The probable error of observations made in this way can best be regarded as one of plate difference, a two-plate difference not being significant, thus a four-negative plate result is not significantly different from a two- or six-negative

plate result. Such a probable error holds so long as the number of plates falls in the intermediate portion of the series, at each extreme end (one or two negative plates or twenty-four or twenty-five negative plates) the error becomes considerable and difficult to calculate. Fortunately in our experiments the number of organisms with the exception of *Heteromita* sp. has been such as to give negative plates within the range of accuracy.

The accuracy of the method has been rigorously tested by a series of experiments given in a earlier paper (6), and the results obtained with *Oicomonas termo* (see p. 335) afford further proof of the efficiency of the technique. This animal shows a two-day periodicity of the active forms which can hardly be ascribed to experimental error.

The number of negative plates are given in Table IV and the actual figures may be obtained by reference to Table III.

TABLE III.—Tables showing the Number of Soil Organisms per gramme of Soil corresponding to the Number of Negative Plates in the various Dilution Series used.

| 1. Dilutions 25–102,400. | | 2. Dilutions 50–204,800 | | 3. Dilutions 50–409,600. | |
|---|---|---|---|---|---|
| No. of sterile plates. | Protozoa per gramme. | No. of sterile plates. | Protozoa per gramme. | No. of sterile plates. | Protozoa per gramme. |
| 1 | 110,000 | 1 | 220,000 | 1 | 420,000 |
| 2 | 59,000 | 2 | 118,000 | 2 | 230,000 |
| 3 | 36,000 | 3 | 72,000 | 3 | 140,000 |
| 4 | 23,000 | 4 | 46,000 | 4 | 95,000 |
| 5 | 16,000 | 5 | 32,000 | 5 | 64,000 |
| 6 | 11,000 | 6 | 22,000 | 6 | 44,000 |
| 7 | 7,600 | 7 | 15,200 | 7 | 30,000 |
| 8 | 5,300 | 8 | 10,600 | 8 | 21,000 |
| 9 | 3,700 | 9 | 7,400 | 9 | 15,000 |
| 10 | 2,600 | 10 | 5,200 | 10 | 10,000 |
| 11 | 1,800 | 11 | 3,600 | 11 | 7,300 |
| 12 | 1,300 | 12 | 2,600 | 12 | 5,100 |
| 13 | 900 | 13 | 1,800 | 13 | 3,600 |
| 14 | 640 | 14 | 1,280 | 14 | 2,600 |
| 15 | 450 | 15 | 900 | 15 | 1,800 |
| 16 | 320 | 16 | 640 | 16 | 1,300 |
| 17 | 230 | 17 | 460 | 17 | 900 |
| 18 | 160 | 18 | 320 | 18 | 640 |
| 19 | 110 | 19 | 220 | 19 | 450 |
| 20 | 79 | 20 | 158 | 20 | 320 |
| 21 | 56 | 21 | 112 | 21 | 220 |
| 22 | 38 | 22 | 76 | 22 | 160 |
| 23 | 25 | 23 | 50 | 23 | 110 |
| 24 | 15 | 24 | 30 | 24 | 77 |
| 25 | 6·8 | 25 | 13·6 | 25 | 51 |
| | | | | 26 | 30 |
| | | | | 27 | 14 |

## Table IV.—Summary of Complete Data (to be used in Conjunction with Table III).

The asterisks against certain bacterial numbers denote that such figures do not conform to the Poisson Series (see p. 340) and have therefore an unknown experimental error. From Sample 1-22 the range of dilutions employed for counting total numbers of Protozoa was 1/25-1/102,400; from Sample 23-123 it was 1/50-1/204,800; and from Sample 124-365 it was 1/50-1/409,600. The range used for estimating the number of cysts throughout the experiment was 1/25-1/102,400.

| Date. | Sample No. | $H_2O$ content. | Rainfall 8″ gauge. | Max. air temp. in °F. | Min. soil temp., 12″ thermometer. | Bacteria in millions per grm. | Protozoa (numbers of negative plates). | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Dimastigamœba gruberi.* | | *Sp. α.* | | *Heteromita sp.* | | *Sp. γ.* | | *Cercomonas sp.* | | *Oicomonas termo.* | |
| | | | | | | | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. |
| 1920. | | | | | | | | | | | | | | | | | | |
| July 5 | 1 | 21·64 | 0·594 | 55·0 | 59·0 | 8·05 | 9 | 10 | 9 | 12 | 3 | 4 | 3 | 4 | 6 | 13 | 3 | 4 |
| ,, 6 | 2 | 22·89 | 0·488 | 61·0 | 57·0 | 30·125 | 13 | 11 | 14 | 18 | 0 | 6 | 1 | 9 | 8 | 13 | 1 | 8 |
| ,, 7 | 3 | 25·10 | 0·158 | 60·1 | 58·0 | 24·825 | 9 | 10 | 12 | 21 | 0 | 0 | 4 | 9 | 14 | 11 | 1 | 2 |
| ,, 8 | 4 | 23·98 | 0·150 | 60·3 | 58·0 | 17·9 | 8 | 10 | 11 | 13 | 0 | 3 | 1 | 13 | 8 | 9 | 0 | 5 |
| ,, 9 | 5 | 23·26 | — | 63·1 | 58·0 | 24·95 | 11 | 10 | 14 | 15 | 3 | 5 | 5 | 7 | 7 | 13 | 5 | 6 |
| ,, 10 | 6 | 21·98 | 0·119 | 61·9 | 57·9 | 9·65 | 10 | 9 | 10 | 25 | 0 | 1 | 4 | 8 | 7 | 11 | 0 | 2 |
| ,, 11 | 7 | 24·54 | 0·115 | 67·1 | 58·3 | 20·75 | 10 | 10 | 10 | 13 | 2 | 3 | 3 | 4 | 5 | 10 | 3 | 3 |
| ,, 12 | 8 | 23·12 | 0·008 | 68·5 | 60·2 | 39·5 | 4 | 4 | 9 | 10 | 0 | 0 | 3 | 11 | 11 | 12 | 2 | 4 |
| ,, 13 | 9 | 21·3 | — | 65·8 | 60·2 | 13·825 | 5 | 10 | 6 | 13 | 2 | 7 | 4 | 7 | 10 | 13 | 3 | 6 |
| ,, 14 | 10 | 20·38 | 0·089 | 66·9 | 59·8 | 12·825 | 6 | 8 | 9 | 11 | 1 | 3 | 5 | 7 | 5 | 11 | 2 | 4 |
| ,, 15 | 11 | 23·09 | 0·005 | 66·9 | 60·0 | 18·375 | 6 | 6 | 5 | 11 | 1 | 1 | 0 | 6 | 7 | 8 | 3 | 2 |
| ,, 16 | 12 | 20·5 | — | 68·1 | 59·3 | 9·875 | 6 | 7 | 7 | 19 | 1 | 4 | 2 | 5 | 4 | 12 | 1 | 13 |
| ,, 17 | 13 | 21·68 | 0·083 | 69·5 | 61·0 | 8·5 | 6 | 10 | 9 | 22 | 2 | 6 | 2 | 13 | 2 | 13 | 4 | 7 |
| ,, 18 | 14 | 19·11 | — | 66·8 | 60·3 | 12·5 | 7 | 9 | 9 | 13 | 0 | 1 | 2 | 2 | 4 | 7 | 4 | 9 |
| ,, 19 | 15 | 19·95 | — | 68·9 | 60·7 | 19·75 | 8 | 7 | 12 | 15 | 0 | 0 | 3 | 2 | 16 | 10 | 1 | 1 |
| ,, 20 | 16 | 21·62 | — | 67·3 | 61·0 | 12·5 | 8 | 9 | 8 | 19 | 1 | 3 | 1 | 5 | 8 | 10 | 2 | 4 |
| ,, 21 | 17 | 20·44 | 0·479 | 66·1 | 61·7 | 12·75 | 11 | 11 | 12 | 10 | 1 | 2 | 1 | 7 | 3 | 13 | 3 | 8 |
| ,, 22 | 18 | 22·18 | 0·004 | 66·9 | 61·8 | 6·5 | 6 | 10 | 8 | 13 | 2 | 5 | 3 | 9 | 7 | 11 | 2 | 10 |
| ,, 23 | 19 | 23·15 | 0·776 | 61·9 | 60·8 | 8·3 | 6 | 7 | 10 | 21 | 1 | 2 | 7 | 10 | 9 | 17 | 3 | 3 |
| ,, 24 | 20 | 22·1 | — | 61·8 | 59·6 | 9·95 | 6 | 9 | 10 | 17 | 4 | 8 | 3 | 9 | 3 | 9 | 3 | 7 |
| ,, 25 | 21 | 23·0 | 0·318 | 61·7 | 58·4 | 9·25 | 7 | 10 | 10 | 18 | 3 | 5 | 8 | 6 | 13 | 12 | 2 | 3 |
| ,, 26 | 22 | 22·74 | 0·016 | 58·9 | 58·5 | 22·062 | 7 | 9 | 10 | 11 | 0 | 5 | 4 | 7 | 1 | 8 | 0 | 5 |
| ,, 27 | 23 | 22·31 | 0·024 | 61·8 | 57·2 | 15·0 | 11 | 11 | 12 | 14 | 4 | 7 | 9 | 10 | 7 | 5 | 4 | 4 |
| ,, 28 | 24 | 22·75 | 0·069 | 61·8 | 57·7 | 16·05 | 13 | 9 | 11 | 13 | 8 | 7 | 1 | 11 | 4 | 12 | 8 | 8 |
| ,, 29 | 25 | 23·02 | 0·099 | 67·0 | 57·2 | 3·7* | 13 | 14 | 7 | 22 | 6 | 9 | 4 | 11 | 10 | 10 | 7 | 8 |
| ,, 30 | 26 | 24·27 | 0·280 | 65·1 | 58·8 | 10·875 | 11 | 9 | 15 | 14 | 4 | 2 | 5 | 3 | 6 | 11 | 4 | 2 |
| ,, 31 | 27 | 23·0 | 0·002 | 69·1 | 58·7 | 14·15 | 9 | 7 | 12 | 12 | 3 | 3 | 4 | 10 | 7 | 6 | 3 | 3 |
| Aug. 1 | 28 | 22·37 | 0·005 | 65·9 | 59·5 | 37·31 | 9 | 8 | 13 | 15 | 4 | 6 | 4 | 9 | 8 | 6 | 3 | 5 |
| ,, 2 | 29 | 21·39 | 0·350 | 60·0 | 59·0 | 17·43 | 8 | 8 | 11 | 12 | 1 | 9 | 5 | 9 | 5 | 10 | 4 | 4 |
| ,, 3 | 30 | 24·18 | — | 65·0 | 59·0 | 16·65 | 11 | 10 | 13 | 21 | 3 | 7 | 6 | 13 | 6 | 11 | 4 | 6 |
| ,, 4 | 31 | 20·14 | 0·158 | 62·2 | 58·8 | 10·0 | 11 | 10 | 9 | 16 | 3 | 8 | 4 | 5 | 4 | 9 | 1 | 6 |
| ,, 5 | 32 | 24·82 | 0·055 | 66·5 | 58·8 | 13·125 | 12 | 11 | 14 | 15 | 3 | 1 | 3 | 9 | 7 | 11 | 3 | 2 |
| ,, 6 | 33 | 22·80 | — | 62·2 | 58·5 | 18·37 | 10 | 8 | 7 | 8 | 3 | 4 | 2 | 8 | 17 | 9 | 3 | 5 |
| ,, 7 | 34 | 22·22 | — | 65·9 | 57·8 | 22·57* | 9 | 10 | 14 | 13 | 2 | 0 | 3 | 4 | 7 | 1 | 3 | 1 |
| ,, 8 | 35 | 18·13 | 0·099 | 71·5 | 58·0 | 8·6 | 7 | 6 | 14 | 13 | 2 | 1 | 4 | 3 | 9 | 14 | 3 | 2 |
| ,, 9 | 36 | 21·91 | 0·024 | 63·1 | 59·7 | 11·5 | 4 | 11 | 8 | 12 | 1 | 6 | 3 | 9 | 3 | 11 | 1 | 8 |
| ,, 10 | 37 | 22·16 | — | 64·6 | 59·0 | 10·0 | 6 | 10 | 13 | 18 | 0 | 3 | 4 | 5 | 4 | 6 | 0 | 3 |
| ,, 11 | 38 | 21·04 | — | 67·8 | 59·0 | 20·0 | 14 | 12 | 13 | 14 | 3 | 6 | 3 | 6 | 7 | 9 | 6 | 6 |
| ,, 12 | 39 | 22·48 | — | 63·9 | 59·5 | 21·3 | 7 | 8 | 10 | 19 | 5 | 1 | 6 | 12 | 5 | 6 | 7 | 2 |
| ,, 13 | 40 | 19·24 | — | 67·6 | 58·7 | 15·25 | 12 | 11 | 4 | 15 | 1 | 8 | 5 | 11 | 6 | 10 | 3 | 11 |
| ,, 14 | 41 | 20·27 | — | 69·2 | 59·0 | 20·05* | 10 | 11 | 15 | 18 | 2 | 3 | 5 | 8 | 8 | 10 | 3 | 6 |
| ,, 15 | 42 | 19·11 | — | 69·3 | 60·3 | 60·625* | 13 | 10 | 14 | 16 | 5 | 3 | 7 | 7 | 8 | 8 | 10 | 5 |
| ,, 16 | 43 | 19·17 | — | 64·1 | 60·8 | 15·07 | 13 | 9 | 10 | 19 | 2 | 1 | 4 | 8 | 5 | 6 | 6 | 12 |
| ,, 17 | 44 | 20·54 | — | 62·2 | 60·7 | — | 8 | 6 | 7 | 18 | 1 | 1 | 1 | 4 | 2 | 10 | 0 | 11 |
| ,, 18 | 45 | 20·00 | 0·520 | 62·2 | 59·8 | 17·5 | 9 | 7 | 5 | 11 | 0 | 6 | 1 | 6 | 9 | 14 | 1 | 4 |
| ,, 19 | 46 | 22·15 | 0·009 | 57·7 | 58·1 | 14·06 | 7 | 10 | 10 | 14 | 1 | 2 | 3 | 11 | 3 | 12 | 2 | 5 |
| ,, 20 | 47 | 21·97 | — | 57·5 | 57·0 | 20·83 | 11 | 8 | 12 | 13 | 4 | 6 | 6 | 6 | 4 | 9 | 2 | 5 |
| ,, 21 | 48 | 21·16 | 0·047 | 57·8 | 56·1 | 13·3 | 6 | 3 | 6 | 11 | 2 | 1 | 3 | 10 | 6 | 3 | 2 | 1 |
| ,, 22 | 49 | 20·32 | — | 59·4 | 56·1 | 18·08* | 8 | 7 | 12 | 13 | 3 | 4 | 5 | 6 | 9 | 8 | 4 | 5 |
| ,, 23 | 50 | 22·11 | — | 61·7 | 55·9 | 10·18 | 10 | 7 | 11 | 19 | 1 | 5 | 4 | 10 | 6 | 12 | 6 | 16 |

| | | |

TABLE IV.—*continued.*

| Date. | Sample No. | $H_2O$ content. | Rainfall 8″ gauge. | Max. air temp. in °F. | Min. soil temp., 12″ thermometer. | Bacteria in millions per grm. | Protozoa (numbers of negative plates). | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Dimastigamœba gruberi.* | | *Sp. α.* | | *Heteromita sp.* | | *Sp. γ.* | | *Cercomonas sp.* | | *Oicomonas termo.* | |
| | | | | | | | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. |
| 1920. | | | | | | | | | | | | | | | | | | |
| Nov. 1 | 120 | 21·39 | 0·004 | 45·8 | 45·1 | 12·85 | 1 | 4 | 6 | 9 | 1 | 0 | 1 | 1 | 1 | 3 | 1 | 2 |
| „ 2 | 121 | 21·35 | — | 42·2 | 44·3 | 31·5* | 2 | 3 | 5 | 14 | 1 | 1 | 10 | 10 | 1 | 2 | 7 | 5 |
| „ 3 | 122 | 22·22 | — | 42·2 | 44·5 | 52·5 | 2 | 4 | 9 | 15 | 1 | 0 | 10 | 9 | 4 | 8 | 9 | 10 |
| „ 4 | 123 | 20·50 | 0·004 | 46·7 | 44·0 | 53·75 | 2 | 4 | 5 | 12 | 1 | 4 | 6 | 8 | 3 | 3 | 3 | 5 |
| „ 5 | 124 | 20·75 | 0·008 | 43·8 | 42·9 | 43·9 | 2 | 6 | 12 | 15 | 2 | 4 | 3 | 10 | 4 | 4 | 6 | 7 |
| „ 6 | 125 | 21·39 | — | 48·3 | 42·5 | 40·45* | 3 | 2 | 8 | 10 | 3 | 1 | 6 | 16 | 4 | 2 | 4 | 3 |
| „ 7 | 126 | 21·52 | 0·004 | 50·9 | 42·5 | 59·41 | 4 | 1 | 6 | 3 | 1 | 3 | 3 | 2 | 3 | 5 | 6 | 5 |
| „ 8 | 127 | 21·00 | — | 49·8 | 43·9 | 59·375 | 4 | 1 | 10 | 12 | 1 | 0 | 0 | 2 | 6 | 6 | 0 | 2 |
| „ 9 | 128 | 21·06 | — | 52·6 | 45·1 | 18·4375 | 3 | 2 | 8 | 14 | 0 | 2 | 2 | 11 | 4 | 2 | 9 | 5 |
| „ 10 | 129 | 20·73 | 0·004 | 53·9 | 46·0 | 52·775 | 7 | 3 | 16 | 11 | 1 | 0 | 4 | 6 | 3 | 3 | 2 | 8 |
| „ 11 | 130 | 20·59 | 0·005 | 50·3 | 46·4 | 67·1875* | 1 | 1 | 8 | 12 | 1 | 1 | 9 | 4 | 2 | 1 | 8 | 5 |
| „ 12 | 131 | 20·80 | 0·035 | 51·7 | 44·2 | 49·875 | 0 | 0 | 8 | 4 | 0 | 0 | 4 | 2 | 1 | 2 | 4 | 2 |
| „ 13 | 132 | 20·36 | 0·080 | 55·1 | 45·0 | 30·375 | 3 | 2 | 4 | 10 | 4 | 1 | 9 | 7 | 2 | 2 | 10 | 2 |
| „ 14 | 133 | 21·28 | 0·375 | 56·4 | 44·7 | 25·9375* | 2 | 5 | 12 | 12 | 1 | 3 | 5 | 10 | 3 | 4 | 7 | 4 |
| „ 15 | 134 | 21·48 | 0·087 | 56·5 | 47·1 | 23·625 | 3 | 3 | 8 | 11 | 3 | 2 | 3 | 3 | 6 | 8 | 4 | 1 |
| „ 16 | 135 | 21·39 | 0·002 | 48·9 | 46·5 | 41·625* | 1 | 1 | 9 | 14 | 4 | 1 | 11 | 6 | 6 | 2 | 6 | 4 |
| „ 17 | 136 | 21·70 | 0·004 | 49·1 | 44·7 | 28·8125* | 3 | 4 | 10 | 12 | 1 | 5 | 7 | 7 | 1 | 3 | 5 | 5 |
| „ 18 | 137 | 21·75 | 0·004 | 47·5 | 42·7 | 25·85 | 4 | 1 | 2 | 11 | 3 | 1 | 15 | 12 | 4 | 2 | 9 | 5 |
| „ 19 | 138 | 20·73 | 0·007 | 48·1 | 43·0 | 25·875 | 3 | 3 | 8 | 12 | 1 | 1 | 3 | 2 | 1 | 0 | 0 | 1 |
| „ 20 | 139 | 22·25 | 0·007 | 38·5 | 43·8 | 58·0* | 1 | 0 | 7 | 8 | 0 | 0 | 7 | 7 | 2 | 0 | 8 | 2 |
| „ 21 | 140 | 22·59 | 0·007 | 40·0 | 41·9 | 25·6 | 0 | 6 | 8 | 17 | 0 | 2 | 1 | 8 | 3 | 3 | 1 | 4 |
| „ 22 | 141 | 21·66 | 0·010 | 42·9 | 40·0 | 53·666 | 5 | 3 | 17 | 16 | 2 | 1 | 12 | 9 | 7 | 4 | 6 | 6 |
| „ 23 | 142 | 20·89 | 0·004 | 45·1 | 39·0 | 56·5* | 2 | 2 | 6 | 7 | 2 | 1 | 7 | 9 | 5 | 3 | 9 | 7 |
| „ 24 | 143 | 21·26 | — | 46·1 | 38·7 | 60·25* | 2 | 2 | 9 | 14 | 1 | 0 | 8 | 6 | 1 | 4 | 1 | 8 |
| „ 25 | 144 | 21·50 | 0·004 | 51·7 | 39·7 | 49·7 | 2 | 2 | 5 | 13 | 1 | 0 | 10 | 9 | 2 | 0 | 3 | 4 |
| „ 26 | 145 | 22·64 | 0·005 | 48·3 | 41·1 | 30·4* | 2 | 3 | 13 | 17 | 1 | 0 | 9 | 6 | 2 | 2 | 5 | 10 |
| „ 27 | 146 | 23·3 | 0·673 | 49·3 | 41·6 | 31·5625* | 1 | 0 | 3 | 10 | 2 | 0 | 13 | 13 | 0 | 1 | 4 | 5 |
| „ 28 | 147 | 21·95 | 0·008 | 47·2 | 43·4 | 44·825* | 4 | 7 | 15 | 10 | 0 | 3 | 1 | 4 | 1 | 3 | 5 | 3 |
| „ 29 | 148 | 22·04 | 0·166 | 47·6 | 43·7 | 52·65 | 4 | 4 | 15 | 11 | 4 | 1 | 9 | 1 | 2 | 3 | 0 | 4 |
| „ 30 | 149 | 23·37 | 0·209 | 49·4 | 44·0 | 21·0625 | 1 | 2 | 9 | 14 | 2 | 1 | 14 | 13 | 3 | 2 | 12 | 9 |
| Dec. 1 | 150 | 23·42 | 0·331 | 47·2 | 43·4 | 30·6 | 0 | 2 | 9 | 14 | 0 | 2 | 3 | 14 | 3 | 3 | 4 | 4 |
| „ 2 | 151 | 22·62 | 0·150 | 53·2 | 42·8 | 45·375 | 5 | 8 | 13 | 16 | 0 | 1 | 8 | 16 | 8 | 8 | 2 | 5 |
| „ 3 | 152 | 23·40 | 0·016 | 54·2 | 43·1 | 49·68 | 8 | 3 | 10 | 12 | 4 | 2 | 13 | 8 | 7 | 2 | 6 | 3 |
| „ 4 | 153 | 22·66 | 0·047 | 44·8 | 42·8 | 12·35* | 0 | 0 | 1 | 8 | 2 | 1 | 11 | 16 | 2 | 2 | 7 | 6 |
| „ 5 | 154 | 21·80 | 0·130 | 43·7 | 42·9 | 44·43* | 6 | 4 | 14 | 13 | 0 | 3 | 4 | 11 | 3 | 3 | 0 | 5 |
| „ 6 | 155 | 22·24 | 0·024 | 39·8 | 42·2 | 8·6 | 2 | 1 | 7 | 2 | 2 | 1 | 7 | 4 | 0 | 1 | 5 | 3 |
| „ 7 | 156 | 22·62 | 0·016 | 39·9 | 41·7 | 54·18 | 1 | 0 | 2 | 4 | 4 | 0 | 5 | 5 | 4 | 4 | 5 | 4 |
| „ 8 | 157 | 23·22 | 0·006 | 36·1 | 41·5 | 49·6 | 3 | 1 | 8 | 7 | 1 | 3 | 6 | 2 | 8 | 5 | 10 | 7 |
| „ 9 | 158 | 21·77 | 0·008 | 38·2 | 40·1 | 41·45* | 3 | 1 | 5 | 7 | 1 | 2 | 5 | 13 | 3 | 5 | 5 | 7 |
| „ 10 | 159 | 22·24 | 0·016 | 37·3 | 39·0 | 25·0* | 3 | 3 | 8 | 8 | 2 | 0 | 10 | 5 | 5 | 2 | 7 | 3 |
| „ 11 | 160 | 22·15 | 0·028 | 34·4 | 38·9 | 17·5 | 2 | 2 | 0 | 9 | 4 | 1 | 8 | 12 | 4 | 0 | 4 | 4 |
| „ 12 | 161 | 23·42 | — | 28·2 | 38·3 | 46·375 | 4 | 2 | 12 | 7 | 1 | 4 | 1 | 4 | 5 | 7 | 5 | 5 |
| „ 13 | 162 | 22·29 | 0·012 | 31·2 | 37·3 | 14·15 | 4 | 1 | 5 | 12 | 7 | 0 | 10 | 8 | 6 | 3 | 5 | 2 |
| „ 14 | 163 | 22·82 | 0·024 | 34·0 | 36·6 | 44·69* | 4 | 1 | 5 | 5 | 0 | 2 | 9 | 10 | 4 | 3 | 6 | 2 |
| „ 15 | 164 | 23·07 | 0·016 | 30·0 | 36·3 | 18·1875 | 8 | 8 | 15 | 11 | 3 | 1 | 9 | 8 | 2 | 4 | 9 | 8 |
| „ 16 | 165 | 23·44 | 0·002 | 33·5 | 36·0 | 25·5* | 4 | 0 | 14 | 17 | 2 | 4 | 4 | 14 | 5 | 3 | 8 | 4 |
| „ 17 | 166 | 21·93 | 0·030 | 35·3 | 35·9 | 24·625 | 2 | 2 | 5 | 10 | 1 | 2 | 12 | 12 | 1 | 3 | 4 | 3 |
| „ 18 | 167 | 22·64 | 0·039 | 37·3 | 36·0 | 55·25* | 2 | 0 | 14 | 12 | 2 | 4 | 6 | 9 | 0 | 3 | 6 | 7 |
| „ 19 | 168 | 23·21 | 0·002 | 36·1 | 37·0 | 46·00* | 2 | 2 | 10 | 16 | 0 | 0 | 8 | 5 | 4 | 1 | 2 | 2 |
| „ 20 | 169 | 22·93 | 0·012 | 43·3 | 37·3 | 21·125 | 3 | 3 | 3 | 5 | 2 | 1 | 10 | 6 | 6 | 6 | 7 | 3 |
| „ 21 | 170 | 22·65 | 0·087 | 45·2 | 38·5 | 31·35* | 3 | 2 | 10 | 9 | 2 | 3 | 4 | 13 | 6 | 4 | 6 | 5 |
| „ 22 | 171 | 23·17 | 0·099 | 43·6 | 38·7 | 17·25 | 2 | 1 | 7 | 16 | 2 | 0 | 15 | 12 | 6 | 1 | 6 | 3 |
| „ 23 | 172 | 23·77 | 0·422 | 48·5 | 38·4 | 29·3125 | 3 | 0 | 17 | 12 | 1 | 1 | 3 | 6 | 4 | 1 | 0 | 3 |
| „ 24 | 173 | 23·85 | 0·052 | 53·1 | 38·2 | 40·575 | 4 | 0 | 16 | 13 | 3 | 0 | 8 | 8 | 5 | 1 | 5 | 5 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| „ | 25 | 174 | 23·17 | — | 50·3 | 41·4 | 35·25 | 1 | 0 | 14 | 13 | 3 | 1 | 12 | 11 | 2 | 3 | 4 | 9 |
| „ | 26 | 175 | 23·37 | 0·032 | 53·0 | 41·3 | 32·2 | 3 | 0 | 16 | 22 | 1 | 1 | 13 | 20 | 1 | 1 | 10 | 5 |
| „ | 27 | 176 | 22·22 | 0·134 | 51·5 | 41·1 | 40·5 | 4 | 2 | 16 | 18 | 0 | 2 | 0 | 10 | 3 | 3 | 2 | 1 |
| „ | 28 | 177 | 22·59 | 0·004 | 52·0 | 42·6 | 45·25 | 0 | 0 | 5 | 14 | 2 | 1 | 16 | 15 | 1 | 1 | 7 | 5 |
| „ | 29 | 178 | 22·44 | 0·259 | 52·3 | 42·8 | 27·69 | 2 | 1 | 14 | 18 | 2 | 1 | 16 | 17 | 6 | 1 | 10 | 2 |
| „ | 30 | 179 | 22·04 | 0·256 | 50·9 | 44·0 | 40·19 | 1 | 1 | 9 | 10 | 3 | 0 | 6 | 2 | 1 | 0 | 5 | 1 |
| „ | 31 | 180 | 24·02 | 0·099 | 53·9 | 43·8 | 52·30* | 3 | 0 | 5 | 15 | 1 | 1 | 9 | 5 | 2 | 1 | 6 | 4 |
| 1921. | | | | | | | | | | | | | | | | | | | |
| Jan. | 1 | 181 | 22·25 | 0·256 | 52·8 | 44·2 | 40·00 | 2 | 1 | 9 | 12 | 1 | 0 | 10 | 14 | 1 | 2 | 5 | 1 |
| „ | 2 | 182 | 23·39 | 0·063 | 51·9 | 45·1 | 35·31 | 2 | 1 | 6 | 13 | 2 | 1 | 8 | 5 | 2 | 1 | 4 | 2 |
| „ | 3 | 183 | 23·70 | 0·002 | 52·1 | 45·0 | 30·05 | 3 | 3 | 10 | 11 | 2 | 3 | 9 | 11 | 2 | 1 | 6 | 4 |
| „ | 4 | 184 | 22·53 | 0·292 | 52·6 | 45·0 | 46·12 | 4 | 0 | 12 | 10 | 2 | 1 | 15 | 13 | 4 | 1 | 10 | 8 |
| „ | 5 | 185 | 23·12 | — | 45·8 | 45·9 | 36·19 | 2 | 1 | 16 | 19 | 0 | 0 | 10 | 16 | 1 | 1 | 7 | 3 |
| „ | 6 | 186 | 22·38 | 0·083 | 47·5 | 44·0 | 38·95 | 3 | 3 | 14 | 18 | 0 | 2 | 13 | 15 | 2 | 6 | 4 | 5 |
| „ | 7 | 187 | 22·68 | 0·488 | 42·9 | 44·1 | 25·13 | 2 | 5 | 15 | 20 | 1 | 0 | 1 | 11 | 1 | 7 | 10 | 9 |
| „ | 8 | 188 | 23·44 | 0·028 | 52·1 | 42·6 | 22·81 | 2 | 4 | 9 | 16 | 2 | 1 | 6 | 17 | 4 | 3 | 3 | 3 |
| „ | 9 | 189 | 23·24 | — | 54·2 | 43·0 | 33·25 | 7 | 3 | 5 | 19 | 0 | 2 | 3 | 18 | 3 | 1 | 6 | 10 |
| „ | 10 | 190 | 23·58 | 0·209 | 53·0 | 44·9 | 14·70 | 2 | 4 | 15 | 10 | 0 | 0 | 3 | 12 | 1 | 1 | 3 | 2 |
| „ | 11 | 191 | 22·36 | 0·087 | 45·7 | 44·0 | 15·94 | 3 | 4 | 11 | 17 | 3 | 1 | 14 | 17 | 6 | 5 | 7 | 3 |
| „ | 12 | 192 | 23·73 | 0·182 | 46·3 | 43·0 | 14·38 | 2 | 2 | 5 | 12 | 2 | 2 | 10 | 16 | 1 | 3 | 4 | 5 |
| „ | 13 | 193 | 23·77 | 0·020 | 42·3 | 42·2 | 29·70 | 2 | 0 | 17 | 13 | 1 | 1 | 12 | 23 | 2 | 1 | 5 | 6 |
| „ | 14 | 194 | 23·19 | — | 38·1 | 41·1 | 23·56 | 1 | 5 | 18 | 19 | 0 | 0 | 17 | 17 | 4 | 6 | 2 | 5 |
| „ | 15 | 195 | 22·33 | 0·002 | 40·8 | 39·2 | 22·88 | 0 | 2 | 13 | 18 | 2 | 0 | 13 | 14 | 4 | 1 | 7 | 3 |
| „ | 16 | 196 | 21·61 | 0·138 | 47·5 | 38·1 | 16·00 | 0 | 2 | 10 | 9 | 2 | 1 | 17 | 15 | 3 | 1 | 3 | 2 |
| „ | 17 | 197 | 23·48 | 0·059 | 51·8 | 39·1 | 22·25 | 1 | 2 | 15 | 17 | 1 | 0 | 4 | 12 | 1 | 2 | 5 | 4 |
| „ | 18 | 198 | 23·09 | 0·170 | 51·9 | 41·8 | 22·38 | 1 | 4 | 14 | 15 | 1 | 4 | 18 | 24 | 1 | 3 | 8 | 4 |
| „ | 19 | 199 | 22·82 | — | 48·2 | 40·1 | 20·95 | 5 | 0 | 16 | 11 | 0 | 0 | 7 | 13 | 6 | 2 | 5 | 6 |
| „ | 20 | 200 | 21·68 | — | 52·2 | 40·1 | 32·13 | 4 | 1 | 17 | 13 | 0 | 1 | 3 | 20 | 3 | 1 | 6 | 2 |
| „ | 21 | 201 | 22·57 | — | 49·2 | 41·6 | 24·40 | 3 | 2 | 7 | 13 | 1 | 1 | 7 | 15 | 1 | 4 | 3 | 5 |
| „ | 22 | 202 | 23·02 | — | 49·3 | 42·3 | 38·05 | 5 | 6 | 14 | 9 | 2 | 1 | 15 | 2 | 4 | 1 | 5 | 2 |
| „ | 23 | 203 | 22·41 | — | 50·7 | 41·3 | 12·69 | 1 | 2 | 7 | 14 | 0 | 0 | 8 | 15 | 0 | 3 | 0 | 3 |
| „ | 24 | 204 | 22·36 | 0·005 | 49·8 | 42·1 | 11·60 | 1 | 3 | 16 | 14 | 0 | 2 | 18 | 16 | 4 | 5 | 7 | 5 |
| „ | 25 | 205 | 22·47 | 0·012 | 50·8 | 43·0 | 22·15 | 5 | 2 | 15 | 12 | 3 | 1 | 18 | 16 | 6 | 2 | 8 | 6 |
| „ | 26 | 206 | 22·86 | 0·012 | 48·9 | 43·5 | 15·50 | 0 | 1 | 19 | 13 | 0 | 1 | 11 | 16 | 4 | 1 | 4 | 7 |
| „ | 27 | 207 | 23·14 | 0·024 | 49·7 | 43·1 | 41·60 | 3 | 1 | 4 | 10 | 1 | 0 | 17 | 20 | 1 | 1 | 7 | 3 |
| „ | 28 | 208 | 22·93 | — | 52·4 | 42·9 | 22·06 | 1 | 1 | 2 | 12 | 0 | 1 | 9 | 9 | 2 | 0 | 2 | 2 |
| „ | 29 | 209 | 22·18 | 0·035 | 49·9 | 44·0 | 23·13 | 2 | 2 | 13 | 12 | 2 | 1 | 4 | 11 | 5 | 2 | 5 | 4 |
| „ | 30 | 210 | 21·30 | 0·114 | 49·5 | 44·8 | 24·81 | 3 | 5 | 14 | 14 | 2 | 3 | 8 | 16 | 3 | 7 | 4 | 6 |
| „ | 31 | 211 | 22·13 | 0·008 | 43·8 | 44·2 | 36·10 | 6 | 8 | 13 | 12 | 3 | 1 | 12 | 9 | 4 | 4 | 5 | 3 |
| Feb. | 1 | 212 | 21·35 | 0·004 | 44·8 | 42·0 | 30·05 | 5 | 0 | 12 | 11 | 1 | 0 | 13 | 18 | 4 | 3 | 11 | 8 |
| „ | 2 | 213 | 22·71 | 0·007 | 43·8 | 40·7 | 30·85 | 4 | 1 | 16 | 19 | 1 | 0 | 19 | 15 | 3 | 1 | 9 | 5 |
| „ | 3 | 214 | 21·80 | 0·002 | 41·2 | 39·8 | 32·75 | 1 | 4 | 11 | 12 | 1 | 1 | 15 | 20 | 1 | 6 | 7 | 5 |
| „ | 4 | 215 | 22·18 | — | 40·8 | 38·5 | 21·45 | 3 | 5 | 16 | 15 | 3 | 7 | 14 | 18 | 2 | 3 | 1 | 5 |
| „ | 5 | 216 | 22·55 | — | 40·1 | 38·7 | 24·10 | 4 | 5 | 20 | 18 | 3 | 3 | 12 | 21 | 2 | 4 | 1 | 1 |
| „ | 6 | 217 | 22·49 | — | 35·9 | 39·0 | 27·62 | 2 | 2 | 19 | 17 | 2 | 2 | 17 | 17 | 4 | 4 | 10 | 4 |
| „ | 7 | 218 | 22·53 | — | 34·8 | 38·2 | 21·95 | 5 | 3 | 7 | 16 | 7 | 3 | 8 | 3 | 5 | 3 | 4 | 1 |
| „ | 8 | 219 | 21·70 | — | 37·8 | 38·1 | 40·50 | 5 | 3 | 11 | 15 | 0 | 0 | 16 | 20 | 4 | 3 | 10 | 5 |
| „ | 9 | 220 | 22·53 | 0·012 | 42·2 | 38·2 | 16·00 | 2 | 3 | 15 | 19 | 1 | 1 | 16 | 21 | 4 | 5 | 8 | 5 |
| „ | 10 | 221 | 22·57 | 0·012 | 43·2 | 38·1 | 27·25 | 5 | 2 | 12 | 12 | 4 | 3 | 15 | 19 | 4 | 1 | 6 | 2 |
| „ | 11 | 222 | 21·14 | 0·020 | 41·3 | 38·2 | 36·60 | 3 | 2 | 14 | 15 | 2 | 5 | 14 | 18 | 0 | 6 | 7 | 7 |
| „ | 12 | 223 | 21·02 | — | 42·5 | 39·0 | 18·80 | 1 | 3 | 17 | 20 | 7 | 1 | 19 | 22 | 8 | 1 | 11 | 7 |
| „ | 13 | 224 | 21·50 | — | 44·7 | 39·2 | 21·20 | 3 | 4 | 8 | 14 | 4 | 4 | 15 | 20 | 5 | 6 | 7 | 8 |
| „ | 14 | 225 | 21·73 | — | 47·9 | 40·0 | 12·65 | 2 | 6 | 11 | 12 | 2 | 2 | 11 | 9 | 2 | 4 | 5 | 2 |
| „ | 15 | 226 | 20·95 | — | 47·1 | 39·5 | 27·40 | 6 | 2 | 6 | 6 | 1 | 1 | 11 | 19 | 4 | 5 | 2 | 4 |
| „ | 16 | 227 | 21·30 | — | 51·2 | 40·1 | 16·88 | 3 | 1 | 10 | 14 | 2 | 3 | 18 | 21 | 4 | 2 | 3 | 6 |
| „ | 17 | 228 | 19·53 | — | 49·8 | 41·7 | 45·38 | 4 | 1 | 17 | 12 | 5 | 5 | 14 | 7 | 3 | 7 | 1 | 5 |
| „ | 18 | 229 | 21·02 | — | 44·9 | 41.0 | 26·55 | 5 | 2 | 13 | 16 | 5 | 3 | 15 | 14 | 4 | 7 | 7 | 9 |
| „ | 19 | 230 | 21·59 | — | 39·6 | 41·0 | 15·90 | 3 | 8 | 8 | 18 | 2 | 7 | 12 | 21 | 4 | 10 | 5 | 11 |
| „ | 20 | 231 | 20·18 | 0·008 | 48·1 | 40·5 | 25·40 | 4 | 3 | 15 | 15 | 0 | 0 | 8 | 1 | 5 | 1 | 1 | 0 |
| „ | 21 | 232 | 19·03 | 0·004 | 50·7 | 39·8 | 12·88 | 1 | 4 | 10 | 5 | 1 | 2 | 14 | 16 | 3 | 3 | 3 | 1 |
| „ | 22 | 233 | 20·02 | 0·008 | 54·2 | 39·0 | 20·60 | 6 | 1 | 19 | 14 | 7 | 4 | 14 | 16 | 7 | 5 | 3 | 5 |
| „ | 23 | 234 | 21·02 | — | 53·1 | 39·9 | 12·38 | 5 | 8 | 11 | 17 | 4 | 2 | 19 | 19 | 8 | 7 | 5 | 7 |
| „ | 24 | 235 | 21·00 | 0·016 | 58·1 | 40·0 | 23·05 | 2 | 5 | 3 | 16 | 4 | 4 | 9 | 17 | 3 | 5 | 4 | 8 |
| „ | 25 | 236 | 21·61 | 0·091 | 48·3 | 40·5 | 15·55 | 6 | 7 | 10 | 17 | 3 | 6 | 15 | 13 | 4 | 8 | 7 | 12 |
| „ | 26 | 237 | 20·91 | 0·004 | 45·4 | 40·3 | 35·50 | 5 | 5 | 15 | 19 | 4 | 5 | 15 | 18 | 4 | 3 | 5 | 5 |
| „ | 27 | 238 | 19·13 | — | 48·0 | 39·1 | 15·95 | 3 | 5 | 6 | 8 | 3 | 3 | 11 | 15 | 5 | 3 | 5 | 2 |
| „ | 28 | 239 | 21·00 | — | 45·9 | 40·0 | 15·35 | 2 | 2 | 16 | 16 | 7 | 0 | 15 | 17 | 3 | 1 | 5 | 8 |
| Mar. | 1 | 240 | 20·09 | 0·092 | 50·0 | 40·1 | 20·56 | 4 | 6 | 6 | 11 | 5 | 2 | 10 | 8 | 5 | 7 | 8 | 4 |
| „ | 2 | 241 | 21·23 | — | 46·5 | 40·2 | 17·19 | 7 | 6 | 20 | 25 | 6 | 7 | 13 | 22 | 7 | 7 | 7 | 9 |

TABLE IV.—*continued.*

| Date. | Sample No. | $H_2O$ content. | Rainfall 8″ gauge. | Max. air temp. in F. | Min. soil temp., 12″ thermometer. | Bacteria in millions per grm. | Protozoa (numbers of negative plates). | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Dimastigamœba gruberi.* | | *Sp. a.* | | *Heteromita sp.* | | *Sp. γ.* | | *Cercomonas sp.* | | *Oicomonas termo.* | |
| | | | | | | | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. | Total. | Cystic. |
| 1921. | | | | | | | | | | | | | | | | | | |
| Mar. 3 | 242 | 20·68 | 0·063 | 48·9 | 39·3 | 14·05 | 2 | 6 | 18 | 14 | 5 | 3 | 14 | 9 | 6 | 4 | 8 | 3 |
| ,, 4 | 243 | 20·95 | — | 51·1 | 40·8 | 26·30 | 4 | 9 | 17 | 13 | 1 | 2 | 11 | 8 | 8 | 4 | 8 | 6 |
| ,, 5 | 244 | 20·70 | 0·002 | 49·1 | 42·0 | 8·88 | 2 | 4 | 7 | 19 | 2 | 1 | 9 | 20 | 2 | 8 | 6 | 3 |
| ,, 6 | 245 | 20·98 | 0·158 | 50·7 | 41·9 | 15·75 | 5 | 8 | 17 | 19 | 3 | 8 | 21 | 22 | 6 | 10 | 12 | 6 |
| ,, 7 | 246 | 20·84 | — | 41·9 | 41·1 | 58·00 | 9 | 5 | 14 | 17 | 1 | 3 | 4 | 25 | 0 | 6 | 1 | 5 |
| ,, 8 | 247 | 20·44 | 0·012 | 45·0 | 39·9 | 14·25 | 3 | 1 | 17 | 22 | 3 | 0 | 11 | 15 | 5 | 4 | 3 | 4 |
| ,, 9 | 248 | 19·37 | — | 51·0 | 40·3 | 26·25 | 6 | 2 | 9 | 5 | 2 | 3 | 13 | 18 | 6 | 2 | 8 | 4 |
| ,, 10 | 249 | 20·79 | — | 55·2 | 41·0 | 16·81 | 9 | 6 | 16 | 14 | 8 | 3 | 15 | 18 | 8 | 5 | 8 | 7 |
| ,, 11 | 250 | 20·48 | — | 55·1 | 42·0 | 36·06 | 8 | 4 | 12 | 8 | 3 | 2 | 12 | 8 | 5 | 6 | 6 | 3 |
| ,, 12 | 251 | 20·38 | 0·035 | 50·2 | 43·0 | 14·25 | 0 | 5 | 12 | 11 | 3 | 2 | 16 | 12 | 4 | 4 | 3 | 3 |
| ,, 13 | 252 | 17·51 | 0·078 | 54·8 | 43·8 | 30·75 | 2 | 1 | 9 | 13 | 3 | 1 | 11 | 9 | 1 | 4 | 6 | 3 |
| ,, 14 | 253 | 20·40 | 0·002 | 53·2 | 44·0 | 16·35 | 1 | 1 | 15 | 16 | 6 | 2 | 15 | 16 | 9 | 6 | 9 | 8 |
| ,, 15 | 254 | 19·85 | — | 53·0 | 43·3 | 22·85 | 4 | 3 | 11 | 12 | 4 | 0 | 14 | 14 | 9 | 4 | 6 | 3 |
| ,, 16 | 255 | 20·20 | 0·051 | 53·9 | 44·7 | 37·80 | 3 | 2 | 12 | 15 | 3 | 0 | 12 | 15 | 4 | 5 | 7 | 7 |
| ,, 17 | 256 | 19·79 | 0·162 | 54·8 | 45·0 | 40·19 | 7 | 5 | 5 | 12 | 2 | 0 | 10 | 11 | 7 | 4 | 5 | 2 |
| ,, 18 | 257 | 21·02 | — | 49·6 | 45·8 | 23·10 | 5 | 6 | 10 | 13 | 2 | 3 | 8 | 3 | 0 | 3 | 1 | 1 |
| ,, 19 | 258 | 20·51 | 0·122 | 50·8 | 44·0 | 16·06 | 2 | 3 | 8 | 5 | 1 | 2 | 10 | 5 | 5 | 5 | 5 | 2 |
| ,, 20 | 259 | 19·39 | 0·028 | 47·2 | 44·0 | 30·55 | 8 | 7 | 4 | 10 | 2 | 3 | 6 | 6 | 3 | 4 | 1 | 2 |
| ,, 21 | 260 | 20·23 | — | 54·1 | 42·8 | 19·60 | 3 | 7 | 10 | 7 | 1 | 1 | 3 | 13 | 3 | 1 | 2 | 3 |
| ,, 22 | 261 | 20·48 | — | 56·8 | 44·7 | 39·19 | 4 | 1 | 11 | 12 | 2 | 1 | 3 | 15 | 5 | 3 | 2 | 6 |
| ,, 23 | 262 | 20·51 | — | 55·1 | 45·8 | 36·69 | 5 | 3 | 11 | 15 | 1 | 0 | 10 | 11 | 3 | 3 | 6 | 4 |
| ,, 24 | 263 | 18·63 | 0·002 | 63·0 | 45·5 | 18·31 | 2 | 5 | 3 | 10 | 2 | 3 | 7 | 13 | 8 | 9 | 2 | 7 |
| ,, 25 | 264 | 19·61 | 0·002 | 58·3 | 45·0 | 18·70 | 3 | 5 | 7 | 11 | 3 | 2 | 8 | 16 | 6 | 6 | 5 | 3 |
| ,, 26 | 265 | 17·90 | 0·001 | 47·8 | 45·0 | 15·25 | 9 | 9 | 12 | 12 | 3 | 7 | 13 | 11 | 9 | 9 | 4 | 6 |
| ,, 27 | 266 | 18·19 | 0·004 | 49·2 | 43·7 | 18·5 | 9 | 5 | 16 | 14 | 2 | 1 | 13 | 12 | 2 | 4 | 4 | 2 |
| ,, 28 | 267 | 17·01 | 0·099 | 51·0 | 43·7 | 13·75 | 2 | 3 | 3 | 12 | 1 | 3 | 6 | 4 | 3 | 6 | 2 | 2 |
| ,, 29 | 268 | 18·04 | 0·051 | 50·0 | 44·0 | 14·5 | 2 | 6 | 13 | 13 | 4 | 2 | 8 | 14 | 4 | 9 | 6 | 5 |
| ,, 30 | 269 | 17·85 | — | 51·1 | 43·4 | 15·625 | 4 | 3 | 4 | 9 | 1 | 2 | 14 | 15 | 2 | 4 | 2 | 5 |
| ,, 31 | 270 | 17·19 | — | 57·8 | 43·9 | 20·15 | 3 | 3 | 2 | 9 | 0 | 0 | 6 | 12 | 4 | 2 | 3 | 1 |
| Apr. 1 | 271 | 20·03 | 0·004 | 61·1 | 45·0 | 28·2 | 2 | 1 | 13 | 9 | 3 | 0 | 13 | 9 | 6 | 6 | 11 | 4 |
| ,, 2 | 272 | 18·61 | — | 64·4 | 45·7 | 18·1875 | 2 | 3 | 13 | 17 | 1 | 1 | 11 | 13 | 5 | 2 | 4 | 7 |
| ,, 3 | 273 | 19·35 | — | 64·0 | 47·0 | 21·125 | 4 | 9 | 16 | 17 | 2 | 0 | 14 | 23 | 2 | 11 | 8 | 6 |
| ,, 4 | 274 | 17·17 | — | 51·1 | 47·2 | 17·10 | 6 | 7 | 1 | 13 | 0 | 1 | 0 | 11 | 4 | 3 | 3 | 4 |
| ,, 5 | 275 | 19·40 | — | 52·7 | 45·8 | 17·7 | 2 | 5 | 13 | 12 | 2 | 1 | 14 | 12 | 2 | 4 | 7 | 2 |
| ,, 6 | 276 | 20·46 | — | 55·1 | 47·0 | 37·625 | 3 | 2 | 15 | 18 | 0 | 0 | 12 | 15 | 1 | 7 | 7 | 7 |
| ,, 7 | 277 | 19·52 | — | 53·2 | 46·4 | 34·2 | 4 | 3 | 10 | 18 | 1 | 0 | 8 | 13 | 2 | 4 | 6 | 4 |
| ,, 8 | 278 | 18·66 | 0·107 | 54·9 | 45·2 | 45·825 | 1 | 3 | 16 | 15 | 0 | 2 | 14 | 18 | 3 | 4 | 5 | 5 |
| ,, 9 | 279 | 16·80 | 0·055 | 51·5 | 45·2 | 20·4 | 0 | 1 | 14 | 21 | 1 | 2 | 10 | 16 | 4 | 5 | 6 | 8 |
| ,, 10 | 280 | 14·06 | — | 63·3 | 45·1 | 13·6875 | 3 | 3 | 14 | 15 | 0 | 0 | 9 | 11 | 4 | 2 | 2 | 2 |
| ,, 11 | 281 | 18·08 | — | 49·9 | 47·0 | 28·1 | 2 | 4 | 9 | 17 | 0 | 6 | 14 | 14 | 5 | 12 | 11 | 8 |
| ,, 12 | 282 | 18·51 | — | 64·9 | 46·9 | 30·35 | 6 | 2 | 13 | 15 | 2 | 1 | 14 | 17 | 7 | 2 | 9 | 5 |
| ,, 13 | 283 | 18·95 | 0·472 | 65·5 | 47·7 | 21·6 | 1 | 5 | 16 | 13 | 0 | 2 | 16 | 13 | 5 | 4 | 12 | 7 |
| ,, 14 | 284 | 17·88 | — | 47·2 | 48·5 | 74·0* | 4 | 3 | 15 | 13 | 1 | 0 | 16 | 19 | 4 | 3 | 5 | 3 |
| ,, 15 | 285 | 16·8 | 0·067 | 40·0 | 45·7 | 61·08 | 6 | 1 | 11 | 18 | 2 | 1 | 11 | 13 | 5 | 3 | 5 | 9 |
| ,, 16 | 286 | 17·96 | 0·248 | 42·9 | 44·1 | 69·1 | 2 | 0 | 14 | 21 | 0 | 0 | 14 | 18 | 4 | 4 | 5 | 3 |
| ,, 17 | 287 | 19·85 | 0·178 | 41·9 | 43·7 | 39·6875 | 1 | 6 | 18 | 22 | 0 | 12 | 18 | 23 | 2 | 9 | 10 | 5 |
| ,, 18 | 288 | 20·48 | 0·016 | 47·1 | 42·7 | 12·55 | 3 | 3 | 18 | 14 | 7 | 0 | 15 | 12 | 4 | 4 | 6 | 8 |
| ,, 19 | 289 | 20·03 | 0·002 | 49·9 | 43·0 | 13·875 | 1 | 3 | 16 | 15 | 2 | 0 | 13 | 17 | 2 | 5 | 9 | 6 |
| ,, 20 | 290 | 19·09 | — | 50·8 | 43·3 | 29·4 | 4 | 5 | 15 | 13 | 1 | 3 | 13 | 15 | 3 | 6 | 6 | 6 |
| ,, 21 | 291 | 18·34 | — | 54·7 | 44·2 | 28·15 | 5 | 4 | 11 | 15 | 1 | 1 | 13 | 21 | 5 | 7 | 7 | 6 |
| ,, 22 | 292 | 18·66 | 0·166 | 54·7 | 46·2 | 20·95 | 1 | 1 | 14 | 20 | 2 | 0 | 12 | 20 | 5 | 2 | 4 | 3 |
| ,, 23 | 293 | 20·88 | 0·103 | 54·4 | 46·9 | 34·8125 | 3 | 2 | 15 | 17 | 0 | 0 | 14 | 19 | 2 | 1 | 6 | 8 |
| ,, 24 | 294 | 19·31 | 0·055 | 48·9 | 46·0 | 24·75 | 3 | 4 | 9 | 12 | 0 | 0 | 12 | 11 | 1 | 6 | 8 | 6 |
| ,, 25 | 295 | 19·4 | — | 53·8 | 45·8 | 20·55 | 5 | 4 | 12 | 21 | 2 | 8 | 17 | 26 | 4 | 11 | 8 | 5 |
| ,, 26 | 296 | 19·95 | — | 52·7 | 46·7 | 22·55 | 4 | 1 | 10 | 9 | 6 | 0 | 18 | 14 | 8 | 2 | 8 | 8 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| " | 28 | 298 | 19·15 | — | 70·2 | 48·7 | 31·55 | 0 | 0 | 10 | 12 | 0 | 0 | 12 | 20 | 2 | 5 | 6 | 9 |
| " | 29 | 299 | 18·64 | — | 65·0 | 49·5 | 22·65 | 2 | 5 | 14 | 16 | 0 | 1 | 9 | 20 | 2 | 7 | 7 | 7 |
| " | 30 | 300 | 18·11 | — | 65·9 | 49·8 | 26·9 | 3 | 1 | 14 | 13 | 0 | 1 | 15 | 16 | 4 | 3 | 5 | 6 |
| May | 1 | 301 | 17·79 | — | 60·0 | 50·1 | 21·0625 | 2 | 1 | 6 | 4 | 0 | 0 | 10 | 6 | 2 | 1 | 4 | 1 |
| " | 2 | 302 | 17·85 | 0·099 | 63·0 | 50·0 | 24·5625 | 2 | 7 | 3 | 0 | 4 | 1 | 5 | 7 | 4 | 1 | 3 | 1 |
| " | 3 | 303 | 16·65 | 0·071 | 45·8 | 50·1 | 32·3 | 5 | 5 | 11 | 15 | 1 | 1 | 16 | 12 | 5 | 8 | 9 | 6 |
| " | 4 | 304 | 16·89 | 0·039 | 45·4 | 48·0 | 26·6 | 4 | 3 | 12 | 15 | 3 | 0 | 15 | 14 | 4 | 5 | 4 | 5 |
| " | 5 | 305 | 16·36 | — | 53·1 | 46·8 | 25·5 | 1 | 6 | 13 | 21 | 0 | 0 | 16 | 24 | 3 | 5 | 8 | 7 |
| " | 6 | 306 | 18·88 | 0·028 | 56·1 | 48·2 | 29·95 | 5 | 4 | 18 | 22 | 0 | 0 | 14 | 19 | 3 | 3 | 3 | 7 |
| " | 7 | 307 | 18·02 | 0·071 | 61·6 | 49·5 | 17·35 | 1 | 2 | 12 | 6 | 1 | 1 | 9 | 3 | 3 | 4 | 5 | 1 |
| " | 8 | 308 | 17·08 | 0·268 | 57·0 | 50·8 | 28·4 | 5 | 4 | 14 | 12 | 0 | 0 | 11 | 9 | 4 | 5 | 6 | 5 |
| " | 9 | 309 | 18·83 | 0·213 | 57·9 | 50·9 | 34·4* | 1 | 0 | 15 | 9 | 4 | 7 | 12 | 9 | 3 | 7 | 8 | 3 |
| " | 10 | 310 | 18·94 | 0·020 | 60·6 | 51·0 | 37·7 | 5 | 2 | 15 | 12 | 2 | 2 | 15 | 13 | 5 | 5 | 8 | 6 |
| " | 11 | 311 | 18·19 | 0·002 | 64·5 | 51·1 | 32·6 | 3 | 3 | 15 | 13 | 1 | 1 | 14 | 12 | 3 | 4 | 6 | 2 |
| " | 12 | 312 | 18·80 | 0·186 | 67·9 | 52·3 | 31·95 | 4 | 3 | 16 | 19 | 0 | 4 | 12 | 21 | 4 | 4 | 5 | 7 |
| " | 13 | 313 | 19·11 | — | 69·9 | 53·5 | 22·85 | 1 | 3 | 13 | 17 | 1 | 2 | 18 | 23 | 6 | 4 | 8 | 10 |
| " | 14 | 314 | 18·42 | 0·153 | 69·2 | 55·0 | 24·9 | 1 | 2 | 14 | 14 | 2 | 1 | 12 | 16 | 0 | 5 | 2 | 5 |
| " | 15 | 315 | 17·46 | 0·020 | 57·9 | 56·0 | 21·5625 | 1 | 4 | 8 | 14 | 0 | 1 | 13 | 12 | 1 | 4 | 2 | 3 |
| " | 16 | 316 | 17·31 | — | 63·0 | 53·8 | 66·4875 | 4 | 1 | 12 | 8 | 9 | 1 | 12 | 5 | 7 | 4 | 7 | 1 |
| " | 17 | 317 | 15·95 | — | 59·5 | 54·1 | 49·125 | 2 | 2 | 14 | 17 | 2 | 1 | 15 | 16 | 5 | 4 | 6 | 6 |
| " | 18 | 318 | 15·30 | — | 62·2 | 54·0 | 23·9 | 2 | 3 | 16 | 23 | 0 | 0 | 20 | 16 | 3 | 3 | 9 | 8 |
| " | 19 | 319 | 15·82 | 0·005 | 66·0 | 55·1 | 12·2 | 1 | 7 | 17 | 21 | 1 | 0 | 14 | 23 | 3 | 4 | 4 | 5 |
| " | 20 | 320 | 15·76 | — | 61·6 | 56·3 | 38·65 | 4 | 4 | 15 | 19 | 3 | 0 | 14 | 18 | 5 | 3 | 8 | 7 |
| " | 21 | 321 | 14·47 | — | 66·8 | 55·9 | 25·66 | 5 | 4 | 11 | 6 | 1 | 1 | 16 | 9 | 5 | 2 | 7 | 4 |
| " | 22 | 322 | 12·72 | — | 69·8 | 56·2 | 22·5 | 5 | 5 | 16 | 14 | 0 | 4 | 15 | 22 | 2 | 7 | 10 | 6 |
| " | 23 | 323 | 14·61 | — | 68·7 | 56·4 | 43·6875 | 7 | 3 | 15 | 8 | 1 | 2 | 12 | 9 | 3 | 2 | 3 | 2 |
| " | 24 | 324 | 16·97 | — | 72·8 | 57·0 | 17·6 | 3 | 3 | 12 | 8 | 3 | 0 | 16 | 11 | 8 | 3 | 8 | 3 |
| " | 25 | 325 | 17·29 | 0·020 | 72·1 | 57·9 | 36·3125 | 4 | 3 | 6 | 7 | 1 | 2 | 6 | 7 | 4 | 2 | 1 | 2 |
| " | 26 | 326 | 17·72 | — | 70·6 | 59·0 | 54·15 | 6 | 4 | 17 | 13 | 3 | 2 | 10 | 19 | 9 | 4 | 8 | 8 |
| " | 27 | 327 | 16·36 | 0·004 | 58·1 | 58·7 | 45·875 | 5 | 6 | 9 | 14 | 2 | 3 | 12 | 19 | 1 | 3 | 0 | 5 |
| " | 28 | 328 | 16·84 | 0·008 | 55·7 | 57·0 | 45·8125 | 3 | 2 | 13 | 17 | 0 | 1 | 11 | 14 | 3 | 2 | 4 | 6 |
| " | 29 | 329 | 14·89 | 0·138 | 57·7 | 56·2 | 36·5 | 3 | 4 | 14 | 14 | 3 | 1 | 16 | 15 | 2 | 5 | 7 | 6 |
| " | 30 | 330 | 17·55 | 0·028 | 64·2 | 56·4 | 17·5 | 2 | 6 | 3 | 12 | 2 | 4 | 7 | 9 | 2 | 4 | 2 | 5 |
| " | 31 | 331 | 16·1 | 0·004 | 63·0 | 57·0 | 48·65 | 4 | 2 | 11 | 17 | 1 | 1 | 16 | 19 | 1 | 1 | 4 | 6 |
| June | 1 | 332 | 16·42 | — | 64·3 | 57·7 | 45·55 | 5 | 4 | 11 | 17 | 1 | 3 | 11 | 19 | 3 | 5 | 2 | 7 |
| " | 2 | 333 | 16·23 | — | 68·3 | 57·5 | 51·35 | 6 | 6 | 8 | 18 | 2 | 1 | 20 | 18 | 0 | 5 | 3 | 6 |
| " | 3 | 334 | 16·25 | 0·083 | 66·8 | 58·7 | 33·15 | 5 | 4 | 17 | 12 | 2 | 0 | 14 | 16 | 7 | 2 | 4 | 1 |
| " | 4 | 335 | 15·0 | — | 56·7 | 57·6 | 49·3125 | 4 | 1 | 15 | 9 | 2 | 0 | 15 | 11 | 2 | 1 | 8 | 2 |
| " | 5 | 336 | 15·69 | — | 63·2 | 56·4 | 14·85 | 1 | 9 | 13 | 13 | 1 | 4 | 6 | 7 | 1 | 5 | 0 | 3 |
| " | 6 | 337 | 15·71 | — | 70·6 | 57·0 | 11·95 | 7 | 3 | 16 | 16 | 5 | 1 | 17 | 12 | 7 | 0 | 5 | 2 |
| " | 7 | 338 | 15·57 | — | 72·8 | 58·0 | 75·55* | 6 | 3 | 12 | 13 | 3 | 0 | 14 | 10 | 3 | 2 | 6 | 4 |
| " | 8 | 339 | 16·57 | — | 69·1 | 58·7 | 14·5 | 2 | 3 | 15 | 17 | 2 | 3 | 13 | 20 | 3 | 3 | 5 | 5 |
| " | 9 | 340 | 15·35 | — | 63·7 | 58·2 | 64·375* | 5 | 2 | 8 | 13 | 1 | 0 | 7 | 15 | 2 | 0 | 1 | 5 |
| " | 10 | 341 | 15·11 | — | 64·8 | 58·0 | 26·15 | 2 | 6 | 10 | 16 | 1 | 1 | 15 | 18 | 1 | 2 | 8 | 3 |
| " | 11 | 342 | 14·83 | — | 64·9 | 58·0 | 34·85 | 4 | 2 | 16 | 16 | 2 | 1 | 12 | 14 | 2 | 3 | 10 | 6 |
| " | 12 | 343 | 13·34 | — | 63·3 | 57·7 | 25·65 | 4 | 7 | 16 | 17 | 0 | 3 | 13 | 18 | 2 | 9 | 8 | 7 |
| " | 13 | 344 | 12·60 | — | 63·3 | 58·5 | 32·75 | 6 | 3 | 14 | 16 | 2 | 1 | 14 | 16 | 6 | 1 | 6 | 6 |
| " | 14 | 345 | 14·22 | — | 66·8 | 57·8 | 14·9375 | 3 | 3 | 9 | 16 | 1 | 0 | 18 | 14 | 2 | 4 | 7 | 3 |
| " | 15 | 346 | 14·51 | — | 76·0 | 58·8 | 27·4 | 7 | 4 | 14 | 17 | 4 | 2 | 13 | 22 | 8 | 4 | 7 | 4 |
| " | 16 | 347 | 12·46 | — | 73·2 | 60·2 | 24·85 | 5 | 8 | 10 | 16 | 5 | 1 | 14 | 15 | 5 | 3 | 9 | 5 |
| " | 17 | 348 | 11·44 | — | 79·5 | 61·0 | 44·875 | 7 | 3 | 15 | 11 | 2 | 0 | 16 | 12 | 5 | 3 | 2 | 1 |
| " | 18 | 349 | 13·30 | — | 61·0 | 61·1 | 39·7 | 3 | 2 | 12 | 12 | 1 | 0 | 13 | 7 | 5 | 2 | 7 | 2 |
| " | 19 | 350 | 10·4 | 0·079 | 57·9 | 58·7 | 34·75 | 2 | 7 | 16 | 23 | 0 | 3 | 7 | 20 | 3 | 4 | 8 | 3 |
| " | 20 | 351 | 13·7 | 0·004 | 59·8 | 58·0 | 30·625 | 6 | 4 | 15 | 13 | 0 | 1 | 17 | 14 | 5 | 6 | 6 | 3 |
| " | 21 | 352 | 13·92 | — | 60·9 | 57·3 | 37·9375 | 8 | 4 | 17 | 24 | 1 | 1 | 19 | 19 | 1 | 3 | 10 | 6 |
| " | 22 | 353 | 13·54 | — | 64·7 | 58·2 | 41·8 | 7 | 4 | 14 | 20 | 0 | 2 | 12 | 23 | 2 | 4 | 4 | 6 |
| " | 23 | 354 | 12·84 | — | 74·7 | 58·7 | 42·375 | 6 | 3 | 13 | 15 | 0 | 1 | 13 | 16 | 4 | 2 | 4 | 4 |
| " | 24 | 355 | 12·44 | — | 81·5 | 60·2 | 56·4* | 7 | 4 | 13 | 13 | 1 | 0 | 17 | 17 | 3 | 0 | 4 | 5 |
| " | 25 | 356 | 12·84 | 0·001 | 82·8 | 63·0 | 34·375 | 4 | 2 | 7 | 22 | 0 | 0 | 13 | 24 | 1 | 0 | 1 | 9 |
| " | 26 | 357 | 11·41 | 0·006 | 71·1 | 63·7 | 65·0 | 10 | 8 | 11 | 6 | 1 | 4 | 9 | 8 | 2 | 6 | 5 | 5 |
| " | 27 | 358 | 12·28 | — | 62·0 | 61·7 | 17·375 | 7 | 7 | 13 | 16 | 2 | 0 | 17 | 19 | 6 | 9 | 2 | 5 |
| " | 28 | 359 | 11·72 | — | 68·8 | 59·8 | 31·7 | 10 | 9 | 13 | 17 | 0 | 6 | 14 | 21 | 5 | 11 | 12 | 10 |
| " | 29 | 360 | 11·55 | — | 67·2 | 61·0 | 35·4 | 8 | 4 | 7 | 6 | 1 | 0 | 9 | 6 | 4 | 1 | 2 | 1 |
| " | 30 | 361 | 10·94 | — | 62·7 | 60·3 | 62·1875 | 3 | 0 | 13 | 21 | 0 | 0 | 14 | 19 | 0 | 0 | 5 | 2 |
| July | 1 | 362 | 10·72 | — | 68·0 | 60·8 | 60·3125 | 7 | 2 | 16 | 14 | 0 | 1 | 15 | 15 | 3 | 1 | 6 | 5 |
| " | 2 | 363 | 11·26 | — | 72·3 | 60·1 | 54·875 | 5 | 1 | 15 | 13 | 0 | 0 | 18 | 19 | 2 | 2 | 9 | 5 |
| " | 3 | 364 | 9·54 | — | 71·5 | 61·3 | 57·0 | 2 | 3 | 14 | 18 | 0 | 2 | 16 | 17 | 2 | 8 | 5 | 10 |
| " | 4 | 365 | 10·26 | — | 61·0 | 60·2 | 17·45 | 6 | 9 | 15 | 16 | 4 | 4 | 15 | 18 | 6 | 6 | 9 | 7 |

## LIST OF REFERENCES.


1. Apstein, C. (1907). "Das Plancton im Colombo-See auf Ceylon," 'Zool. Jahrb., Abt. f. Syst.,' vol. 25.
2. Cash, J. (1905). "British Fresh-water Rhizopoda and Heliozoa," 'Ray Society, Rhizopoda,' vol. 1, Pt. I.
3. Crump, L. M. (1920). "Numbers of Protozoa in certain Rothamsted Soils," 'Journ. Agric. Sci.,' vol. 10.
4. Cunningham, A. (1915). "Studies on Soil Protozoa," *Ibid.*, vol. 7, Pt. I.
5. Cunningham, A., and Lohnis, F. (1913). "Studies on Soil Protozoa," 'Centrlb. f. Bakt.,' Ab. II, vol. 39.
6. Cutler, D. W. (1919). "Observations on Soil Protozoa," 'Journ. Agric. Sci.,' vol. 9.
7. *Idem* (1920). "A Method for Estimating the Number of Active Protozoa in the Soil," *Ibid.*, vol. 10.
8. Cutler, D. W., and Crump, L. M. (1920). "Daily Periodicity in the Numbers of Active Soil Flagellates, with a brief note on the Relation of Trophic Amœbæ and Bacterial Numbers," 'Annals of Applied Biology,' vol. 7.
9. Dangeard, P. A. (1900). "Étude de la Karyokinèse chez l'Amœba hyalina, sp. nov.," 'Le Botaniste,' 7th sér.
10. *Idem* (1910). "Études sur le Développement et la Structure des Organismes Inferieurs," *Ibid.*, 11th sér.
11. Delf, E. M. (1915). "The Algal Vegetation of some Ponds on Hampstead Heath," 'New Phytologist,' vol. 14.
12. Dobell, C. C. (1908). "The Structure and Life-History of *Copromonas subtilis*," 'Quart. Journ. Micro. Sci.,' vol. 52.
13. *Idem* (1913). "Observations on the Life-History of Cienkowski's Arachnula," 'Arch. f. Prot.,' vol. 31.
14. *Idem* (1914). "Cytological Studies of three Species of Amœba," *Ibid.*, vol. 34.

14a. Engberding, D. (1909). "Vergleichende Untersuchungen über die Bakterienzahl im Ackerboden," 'Centrl. f. Bakt. II,' vol. 23.

14b. Francé, R. H. (1921). "Das Edaphon."

15. Fritsch, F. E. and Rich, F. (1913). "A Four Years' Observation of a Fresh-water Pond," 'Annales de Biologie Lacustre,' vol. 6.
16. Gläser, H. (1912). "Untersuchungen über die Teilung einiger Amöben," 'Arch. f. Protistenk.,' vol. 25.
17. Goodey, T. (1916). "Further Observations on Protozoa in Relation to Soil Bacteria," 'Proc. Roy. Soc.,' B, vol. 89.

17A. Goodey, T. (1916). "Observations on the Cytology of Flagellates and Amœbæ obtained from Old Stored Soils," 'Proc. Zool. Soc.,' No. 21.

18. Hartmann, M., and Nägler, K. (1908). "Copulation bei *Amœba diploidea*, n. sp., mit Selbständigbleiben der Gametenkerne während des ganzen Lebenszyklus," 'Sitz.-Ber. Gesellsch. Naturf. Freunde,' Berlin. No. 4.

19. Herdman, W. A. (1922). "Spolia Runiana V.," 'Journ. Linn. Soc." (Botany), vol. 46.

20. Hodgetts, W. J. (1921). "The Periodicity of Fresh-water Algæ," 'New Phytologist,' vol. 20.

21. Keen, B. A., and Russell, E. J. (1921). "Factors determining Soil Temperature," 'Journ. Agric. Sci.,' vol. 11.

22. Klebs, G. (1893). "Flagellatenstudien," 'Zeit. Wissensch. Zool.,' vol. 55.

22A. Kofoid, C. A. (1903 and 1908). "The Plankton of the Illinois River," 'Bull. Illinois State Lab. Nat. Hist.'

23. Lemmermann, E. (1914). "Flagellatæ" in 'Die Süsswasser-Flora Deutschlands.'

24. Leidy (1879). "Fresh-water Rhizopods of North America," 'Report of the United States Geological Survey.'

25. Martin, C. H. (1912). "A Note on the Protozoa from Sick Soils, with some Account of the Life-Cycle of a Flagellate Monad," 'Proc. Roy. Soc.,' B, vol. 85.

26. *Idem* (1913). "Some Remarks on the Behaviour of the Kinetonucleus in the Division of Flagellates, with a Note on *Prowazekia terricola*, a new Flagellate from Sick Soil," 'Zool. Anz.,' vol. 41.

27. Martin, C. H., and Lewin, K. (1914). "Some Notes on Soil Protozoa," 'Phil. Trans.,' B, vol. 205.

28. *Idem* (1915). "Notes on Some Methods for the Examination of Soil Protozoa," 'Journ. Agric. Sci.,' vol. 7.

29. Nägler, K. (1909). "Entwicklungsgeschichtliche Studien über Amöben," 'Arch. f. Protistenk.,' vol. 15.

29A. Penard, E. (1902). "Faune Rhizopodique du Bassin du Léman, Geneva."

30. Puschkarew, B. M. (1913). "Über die Verbreitung der Süsswasser-Protozoen durch die Luft," 'Arch. f. Protistenk.,' vol. 28.

31. (1909.) "Royal Commission on Sewage Disposal," Dorking Exp. Station.

32. Russell, E. J. (1915). "Soil Protozoa and Soil Bacteria," 'Proc. Roy. Soc.,' B, vol. 89.

33. *Idem* (1921). "Soil Conditions and Plant Growth," Fourth Edit. Rothamsted Monographs on Agricultural Science; Longmans, Green and Co., London.

34. Russell, E. J., and Appleyard, A. (1915). "The Atmosphere of the Soil," 'Journ. Agric. Sci.,' vol 7.

35. West, W., and West, G. S. (1912). "On the Periodicity of the Phytoplankton of some British Lakes," 'Journ. Linn. Soc. Botany,' vol. 40.

36. Wilson, C. W. (1916). "On the Life-History of a Soil Amœba," 'Univ. Calif. Publ. Zool.,' vol. 16.

37. Woodcock, H. M. (1916). "Observations on Coprozoic Flagellates, together with a Suggestion as to the Significance of the Kinetonucleus in the Binucleata," 'Phil. Trans.,' B, vol. 207.

# III. *The Influence of Rainfall on the Yield of Wheat at Rothamsted.*

*By* R. A. Fisher, *M.A., Head of the Statistical Department, Rothamsted Experimental Station, Harpenden, Fellow of Gonville and Caius College.*

*Communicated by* Sir John Russell, *F.R.S.*

(Received September 26,—Read November 8, 1923.)

Contents.

| | | Page. |
|---|---|---|
| 1. | General Problem of Evaluating Effects of Weather on Crops | 89 |
| 2. | Effects of Paucity of Crop Data | 90 |
| 3. | Analysis of the Season | 95 |
| 4. | Correlation of Residuals | 99 |
| 5. | Rothamsted Rain and Wheat Data | 108 |
| 6. | Rain Distribution Values | 113 |
| 7. | Correlation of same | 120 |
| 8. | Regression of Yield on Distribution Values and on Rainfall at Different Seasons | 122 |
| 9. | Discussion of Diagrams | 123 |
| 10. | Value of Rainfall Regressions as Prediction Formula | 132 |
| 11. | Comparison with Previous Results | 136 |
| 12. | Summary | 141 |
| 13. | References | 141 |

## 1. *General Problem of Evaluating the Effects of Weather on Crops.*

At the present time very little can be claimed to be known as to the effects of weather upon farm crops. The obscurity of the subject, in spite of its immense importance to a great national industry, may be ascribed partly to the inherent complexity of the problem which it presents, and more especially to the lack of quantitative data obtained either under experimental or under industrial conditions, by the study of which accurate knowledge alone can be acquired. Of the industrial applications of such knowledge it is unnecessary to speak here in detail. It is sufficient to indicate that the present system of Life Insurance, which safeguards the economic stability of many thousands of families, and occupies the activities of many of the greatest financial corporations, was made possible by the studies of statistics of human mortality by the mathematicians of the eighteenth and nineteenth centuries, and that on the basis of adequate knowledge similar economic stability with its attendant security of capital should be within the reach of the industrial farmer.

The inherent complexity of the relationships which it is sought to elucidate, between the yields of farm crops, and the previous weather which largely controls those yields, arises primarily from the complexity of the problem of specifying the weather itself. Meteorologists have, however, gradually devised a number of instrumental observations, which although far from specifying the total environment of the growing plant, as understood by the plant physiologist, do nevertheless give a sufficiently detailed account of the general environmental conditions of the growing crop, in so far as these vary from season to season. It is probable indeed that almost all the weather influences to which crop variations are due could be expressed in terms of the instrumental observations of a modern meteorological station. The actual difficulty of calculating the crop variations from given instrumental records is, however, immense ; only an attack of the most preliminary kind upon the general problem can be attempted in this paper. The complete aim of agricultural meteorology should, however, be emphasised, for it is only by its substantial achievement that other causes of crop variation can be freed from much obscurity.

A valuable example of an investigation in which this aim was clearly held in view is to be found in A. Walter's study in 1910 of the effects of weather upon the Sugar crop in Mauritius (17). Walter was able to give a very complete account of the meteorological causes affecting the yield of sugar, to such an extent that this author considers that under uniform estate management and methods of cultivation the yield may be predicted from meteorological data only with a standard error of the order of 1 per cent. The number of harvests available for this enquiry seems, however, to have been only 13, and more recent investigations in statistical theory make it now possible to realise, more clearly than was at the time possible, some of the dangers which arise from paucity of crop data.

### 2. *Effects of paucity of Crop Data.*

The biometrical investigations for which the method of multiple correlation was developed, differ from such agricultural studies as the present in two main particulars. In the first place the number of individuals measured was taken to be large, of the order of 1,000, or at least of some hundreds, and the special problems of distribution which arise in small samples have only more recently begun to receive attention. In the second place the number of measurements taken of each individual, or the number of *variates*, was generally small, and in all cases far smaller than the number of meteorological elements which may plausibly be regarded as affecting the crop. The sequence of weather to which crop variations may be ascribed extends over a year, or even more, and consists in this country of a series of abrupt, relatively violent and transient spells, each of which has its influence on the crop. If we wished to analyse the sequence no more closely than by monthly averages, we should still have 12 values for rainfall, and 12 more for maximum and minimum temperature, dew point, grass minimum, solar maximum and soil temperatures, nor would it be unreasonable to include some such measure of insolation as is given by "Hours of Bright Sunshine," and averages for the direction and force

of the wind. The number of meteorological elements might be made to exceed even the longest series of crop records available, for the Rothamsted wheat records provide, after necessary deductions, only 60 yields. Consequently, if the computer were provided not with yield data at all, but with an equal number of values composed at random, he would still be able to express them with perfect accuracy in terms of the weather records, for the number of unknowns available would exceed the number of equations for them to satisfy.

It is of more practical importance that even when we have selected a number of meteorological variables which is less than the number of crops recorded, a strong semblance of dependence may be produced, even when fictitious data, unrelated to the weather, are substituted for the true crops. For let there be $n$ values of the dependent variate $y$ and $n$ each of the $q$ meteorological variates $x_1, x_2, \ldots, x_q$. If now we obtain a partial regression equation

$$Y = \bar{y} + S(c_k(x_k - \bar{x}_k)),$$

showing the apparent dependence of $y$ upon $x_1, \ldots, x_q$, then the correlation between $y$ and Y, namely R, will be the multiple correlation of $y$ with $x_1, \ldots, x_q$, and will be greater than the correlation of $y$ with any other linear function of $x_1, \ldots, x_q$. Since R is necessarily positive (or zero), as was noticed by Hooker (10, p. 7), there will be generally an appearance of correlation produced in this way, even if no such relation really exists. This effect becomes particularly marked if the number in the sample is small.

When there is no real correlation it is possible to determine *a priori* the distribution of the multiple correlation of any quantity $y$ with any $q$ variates $x_1, x_2, \ldots, x_q$. There are many ways of approaching the solution, of which perhaps the most enlightening is to regard the deviations $y - \bar{y}$ as components determining the direction of a radius vector OP in $n$ dimensions. Then $q$ other directions are specified by the $q$ meteorological variates, and any linear regression formula is represented by some line through the origin in the space of $q$ dimensions, specified by these $q$ directions. The multiple correlation, R, is then the cosine of the angle which the line OP makes with the space of $q$ dimensions. If the variate $y$ is unrelated to the variates $x_1, \ldots, x_q$, the line OP may be regarded as drawn at random through O, in the space of $n$ dimensions; using this fact it may be shown without difficulty that the chance that $R^2$ falls into the elementary range $dR^2$ is

$$df = \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-q-3}{2}\right)!\left(\frac{q-2}{2}\right)!}(R^2)^{\frac{q-2}{2}}(1-R^2)^{\frac{n-q-3}{2}}d(R^2). \quad \ldots\ldots (1)$$

It will be noted that the distribution is independent of any correlations which may exist between the meteorological variates, provided no one of them can be exactly

calculated from the others. In the case $q = 1$, when only one independent variate is used, the formula reduces to

$$df = \frac{2.\frac{n-3}{2}!}{\frac{n-4}{2}!\sqrt{\pi}}(1-R^2)^{\frac{n-4}{2}}dR,$$

which is known (FISHER (4), 1915) to be the distribution of the correlation (taken positive) derived from a sample of $n$, between two uncorrelated variates. It may also be seen that when $q = n - 1$. R is necessarily equal to $\pm 1$.

The mean value of $R^2$ is

$$\overline{R^2} = \frac{q}{n-1},$$

and since

$$S(Y-\bar{y})^2 = R^2 S(y-\bar{y})^2,$$

we may say that in the absence of correlation, the variance of $y$ (the square of its standard deviation) will be distributed on the average equally between the $q$ degrees of freedom of the regression formula, and the $(n - q - 1)$ degrees of freedom in which $y$ departs from the regression formula. This view of the matter enables us to follow the above conclusions to some extent into the more general case in which $y$ is really in some degree correlated with $x_1, \ldots, x_q$. For if $y$ be imagined as made up of two parts, $y_0 + y'$, such that $y_0$ is wholly determined as a linear function by the independent variates and $y'$ is wholly uncorrelated with them, then $y_0$ and $y'$ must be mutually uncorrelated, so that each will contribute a certain percentage to the variance of $y$. Let the fraction contributed by $y_0$ be A, then the regression formula found by correlating a sample of $y'$ with the independent variates will differ from that found by correlating $y$, merely by $y_0$ which is by hypothesis a linear function of $x_1, \ldots, x_q$. Consequently the average proportion of the variance in the regression formula for $y$ will be

$$\overline{R^2} = A + \frac{q}{n-1}(1-A),$$

or

$$1 - \overline{R^2} = (1-A)\frac{n-q-1}{n-1}.$$

If the sample be increased indefinitely $(n \to \infty)$ then the limiting value of $R^2$ is A. An estimate of A from a finite sample, free from the positive bias of $R^2$, is obtained by taking

$$1 - A_1 = \frac{n-1}{n-q-1}(1-R^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (II)$$

A complete discussion of the errors of random sampling in multiple regression requires a knowledge of the frequency distribution of R, or of $A_1$, calculated by equation (II), in the general case when A is not zero; equation I gives this distribution only when

$A = 0$. It provides a means of testing if an observed value of R differs significantly from zero, but not for testing in general, whether one value of $R^2$ is significantly greater than another.

From equation (1) it appears that the probability that R, obtained from a random sample of $n$ observations, should exceed any specified value, is, if $q$ is even,

$$P = (1 - R^2)^{\frac{1}{2}(n-q-1)} \left\{ 1 + \frac{n-q-1}{2} R^2 + \frac{(n-q-1)(n-q+1)}{2.4} R^4 + \ldots \frac{(n-q-1)\ldots(n-5)}{2.4\ldots(q-2)} R^{q-2} \right\},$$

and, if $q$ is odd,

$$P = \frac{2}{\sqrt{\pi}} \frac{\left(\frac{n-q-2}{2}\right)!}{\left(\frac{n-q-3}{2}\right)!} \int_{\sqrt{\frac{R^2}{1-R^2}}}^{\infty} \frac{dz}{(1+z^2)^{\frac{1}{2}(n-q)}} + \frac{2}{\sqrt{\pi}} \frac{\left(\frac{n-q-2}{2}\right)!}{\left(\frac{n-q-3}{2}\right)!} R (1 - R^2)^{\frac{1}{2}(n-q-1)} \left\{ 1 + \frac{n-q}{3} R^2 + \ldots \frac{(n-q)(n-q+2)\ldots(n-5)}{3.5\ldots(q-2)} R^{q-3} \right\}.$$

The analogy of these formulæ with those giving the probability, P, that $\chi^2$, the Pearsonian test of goodness of fit, should exceed any specified value is obvious. It will be noticed that in the second formula the probability integral of the normal curve is replaced by that of the Type VII curves which has been tabulated by "Student" (16, 1917). When $q$ is small, very few of the terms of the series are involved, thus for $q = 6$, or 7, the series terminates with the term in $R^4$.

The effect of increasing $q$, the number of variates used, is to increase somewhat rapidly the probability that R should exceed any specified value, even for an independent variate wholly uncorrelated with those used to predict it; this increase may be exemplified in the following table, where we have taken $n = 13$, $q = 4$, 6, 8 and calculated the chance of R exceeding 0·5, 0·6, 0·7, 0·8 and 0·9.

Table I.—$n = 13$.

| $q$ | R = 0·5 | 0·6 | 0·7 | 0·8 | 0·9 |
|---|---|---|---|---|---|
| 4 | 0·6328 | 0·4094 | 0·2002 | 0·0598 | 0·0055 |
| 6 | 0·8965 | 0·7491 | 0·5187 | 0·2509 | 0·0505 |
| 8 | 0·9844 | 0·9402 | 0·8248 | 0·5906 | 0·2124 |

It will be seen that the chance of obtaining a value $R > 0{\cdot}9$ is only one in 180 for $q = 4$; it rises to one in twenty for $q = 6$, and about one in four for $q = 8$. For $q = 8$, therefore, an observed multiple correlation $0{\cdot}9$ cannot be regarded as significant, that is as convincing evidence that the "crop" is in any way influenced by the meteorological variates. The only value, in fact, in the above table which could be regarded as definitely significant is the value $0{\cdot}9$ for $q = 4$, while for $q = 6$ this is suggestive only of real influence.

A still more insidious source of illusory high correlations lies in the fact that the particular variates chosen for correlation with the crop figures, are often chosen *because* they appear in fact to be associated with the crop. It is a common practice, as a preliminary to the study of weather correlations, to search for the so-called critical periods by such methods as the following. The years are arranged in order of crop yield, and a number of meteorological values are plotted on a chart in this order; those meteorological values which show an apparent trend upwards or downwards, are then picked out, and used to construct a weather formula by which the crop may be predicted. Such methods are especially deceptive when the series of years is short (about 20 or 30), for if such a process were carried out thoroughly we should not have merely the random sampling distribution of R, which as we have seen is capable of yielding sufficiently high values from uncorrelated material, but we should be choosing that set of $q$ variates, out of a larger number $p$, which gave the highest value of R.

The effect of such a process, if dummy data were substituted for the actual crop records, could be accurately foretold by the solution of the following problem. A line OP is drawn through a point O, at random in $n$ dimensions, $p$ directions (where $p$ may exceed $n$) are also chosen at random, and of them $q$ ($q < n$) are selected so that the space of $q$ dimensions through these makes the least possible angle, $\theta$, with OP. Find the sampling distribution of $\theta$ (or, of $R = \cos\theta$).

I can put forward no general solution of this problem; the case $p = q$ has been solved above. The additional advantage conferred by a choice of variates may be clearly seen from the case $q = 1$. For

$$\int_0^R \frac{2 \cdot \frac{n-3}{2}!}{\sqrt{\pi} \cdot \frac{n-4}{2}!} (1 - R^2)^{\frac{n-4}{2}} dR$$

represents the chance that for a single uncorrelated variate chosen at random, the correlation in a sample shall not exceed R; hence it follows that the chance that for none of $p$ uncorrelated variates chosen at random, does the correlation exceed R, must be

$$\left\{ \int_0^R \frac{2 \cdot \frac{n-3}{2}!}{\sqrt{\pi} \cdot \frac{n-4}{2}!} (1 - R^2)^{\frac{n-4}{2}} dR \right\}^p;$$

and this is the chance that the highest correlation of $p$ shall not exceed R.

If a standard of significance be chosen such that it will be exceeded in random samples once in 20 times, the effect of having a choice of $p$ variates will increase the chance of exceeding any assigned rare value nearly $p$ times. Thus for $n = 13$, $0{\cdot}6$ will not be judged significant, even for a single variate correlated if $p$ exceed 2, $0{\cdot}7$ ceases to be significant if $p > 7$, and $0{\cdot}8$ when $p$ exceeds about 54. When we consider that $p$ may be very great, since not only is the number of meteorological elements available for selection large, but also since many writers allow themselves to use complicated functions of the instrumental data, involving adjustable weights, special conventions of sign, and allowances for periods judged to be unusual in their effects, so that these artificial meteorological data may be calculated in an enormous variety of ways, it is clear that the conclusions we have drawn as to the dangers of applying multiple correlation formulæ to small samples, are very much to be emphasised when a choice is made as to what meteorological elements to correlate.

In view of the foregoing facts it would seem worth while to lay down the following conditions for arriving at unprejudiced results :—

(i) The meteorological variates to be employed must be chosen without reference to the actual crop record.

(ii) If multiple variates are to be used allowance must be made for the positive bias of $R^2$.

(iii) Relationships of a complicated character should be sought only when long series of crop data are available.

### 3. *The Analysis of the Season.*

By the Season is meant the whole sequence of weather directly or indirectly influencing the crop from its inception to the time the produce is weighed. In studying the influence of the rain on the wheat crop, we have chosen a period of a year ending on the 31st of August of the year in which the crop was harvested. To obviate the irregularity of the calendar, this year was taken to be of 366 days, so that it commenced on either the 31st of August or the 1st of September prior to the sowing of the seed. Taking into consideration a single element, namely Rain, only, this sequence is in our climate sufficiently complex. The general distribution of rain through the year could, indeed, be roughly represented by dividing the period up into say 6 sections and recording the total rainfall in each period. If the rainfall were approximately evenly distributed in successive days or weeks, this method might indeed give a fairly accurate picture of the sequence of rainfall. But notoriously the rain is in fact often concentrated in short spells, with rainless periods intervening, consequently if March and April formed one section, a great part of the rain ascribed to that section might fall early in March in one year, and late in April in another, and it would be impossible to regard such falls as equivalent

in their effects on the crop. Whereas a late April fall might well be nearly equivalent to one occurring early in May.

This consideration suggests that a finer subdivision of the year is necessary; that months or even weeks should be treated separately. From the agricultural and meteorological point of view there is everything to be said for such subdivision; it raises, however, new mathematical difficulties; for whereas the evaluation of determinants of 6 rows and columns, or the solution of simultaneous linear equations in 6 unknowns is sufficiently rapid, disproportionate labour is involved in increasing the 6 to 12, and if 52 unknowns were attempted the labour involved would become fantastic.

A consideration which has been still more influential in framing our method is that even if we were to calculate the 52 partial regression coefficients showing the average effect in bushels per acre of an extra inch of rainfall, for each week of the year, such a calculation would leave out of consideration the all-important fact that this effect may be expected to change *continuously* during the year. The true partial regression coefficients for neighbouring periods must be relatively alike. The method of partial correlation as hitherto developed, takes no account of the serial character of our weather variates; after calculating the partial regression coefficients, we should still be far from the facts, if we did not smooth the series so obtained by a continuous curve, which should average out the independent errors in the values obtained for the successive weeks.

Disregarding, then, both the arithmetical and the statistical difficulties, which a direct attack on the problem would encounter, we may recognise that whereas with $q$ subdivisions of the year, the linear regression equations of the wheat crop upon the rainfall would be of the form

$$\bar{w} = c + a_1 r_1 + a_2 r_2 + \ldots + a_q r_q$$

where $r_1, r_2, \ldots, r_q$ are the quantities of rain in the several intervals of time, and $a_1, \ldots a_q$ are the regression coefficients, so if infinitely small subdivisions of time were taken, we should replace the linear regression function by a *regression integral* of the form

$$\bar{w} = c + \int_0^T a r \, dt, \qquad \text{(III)}$$

where $r\,dt$ is the rain falling in the element of time $dt$; the integral is taken over the whole period concerned, and $a$ is a *continuous* function of the time $t$, which it is our object to evaluate from the statistical data.

It will be seen that by proceeding to the limit all artificiality has been eliminated from the quantity $a$, which now represents an objective physical quantity, namely the average benefit to the crop in bushels per acre per inch of rain, falling in the time-element considered. This, of course, is more than even a daily record of rain can tell us, but owing to the relatively slow changes in the functions $a$, we shall find it sufficient to divide the 366 day year into 61 equal periods of 6 days each.

It should be noted that corresponding to the quadratic terms of a regression formula the independent variates of which form a continuous series, *i.e.*, to

$$\mathrm{S}\,(b_{pq}r_p r_q),$$

we have the double regression integral

$$\int_0^{\mathrm{T}}\int_0^{\mathrm{T}} b\,(t,\,t')\,r_t r_{t'}\,dt\,dt',$$

where $b$ is a continuous function of the two epochs $t$ and $t'$, just as $a$ is a continuous function of $t$ only.

The concept of the regression integral, involving a regression function varying continuously with the time, is not only of service in displaying in a precise mathematical form the nature of the relationship which is to be investigated, but when the functions $a$ and $b$ vary with the time relatively slowly, it suggests a statistical procedure by which values of these functions may be obtained from the data. That the values must change relatively slowly is apparent from the fact that the state of advancement of the crop, as indicated either by the time of harvest, or by phenological observations on common weeds, often varies relatively to the Calendar date by as much as a fortnight.

If, then, $\mathrm{T}_0$, $\mathrm{T}_1$, $\mathrm{T}_2$ ... be a series of orthogonal functions of the time, such that

$$\int_0^{\mathrm{T}} \mathrm{T}_r\mathrm{T}_s\,dt = 0 \qquad (r \neq s)$$

$$\int_0^{\mathrm{T}} \mathrm{T}_r^2\,dt = 1,$$

we may express the rate of rainfall at any epoch, in the form

$$r = \rho_0\mathrm{T}_0 + \rho_1\mathrm{T}_1 + \rho_2\mathrm{T}_2 + \ldots,$$

where

$$\rho_s = \int_0^{\mathrm{T}} r\mathrm{T}_s\,dt;$$

also we may express the regression function in the form

$$a = \alpha_0\mathrm{T}_0 + \alpha_1\mathrm{T}_1 + \alpha_2\mathrm{T}_2 + \ldots, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(IV)}$$

where

$$\alpha_s = \int_0^{\mathrm{T}} a\mathrm{T}_s\,dt,$$

noting in the latter case that relatively few terms of the expansion will be required, if $a$ is a slowly-varying quantity.

Then we shall have from equation (III)

$$\overline{w} = c + \int_0^T ar\,dt = c + \alpha_0\rho_0 + \alpha_1\rho_1 + \alpha_2\rho_2 + \ldots .$$

Now the values of $\rho_0$, $\rho_1$, $\rho_2$, .... may be obtained for each year from the rain record, and by correlating them with the crop data, we can obtain the values of $\alpha_0$, $\alpha_1$, $\alpha_2$, ..., as partial regression coefficients of the crop on the coefficients of the rainfall distribution; then from (IV) we can evaluate the expansion of $a$ to as many terms as we have obtained. It is easy to verify that in the same manner the double regression integral may be expressed as a quadratic regression formula for the crop in terms of the rainfall distribution values.

The advantage of this indirect method of attack lies not only in the facts that for a fine division of the year, the arithmetical difficulties of a direct attack are insurmountable, and that no existing crop record is sufficiently long to justify the calculation from it of 50 or 60 independent regression coefficients, for the probable error of such determinations depends on the excess of the number of observations over that of the coefficients evaluated (9), but also in the increased accuracy attained through utilising our knowledge that the regression function must in fact vary relatively slowly. This advantage is analogous to that obtained by the author in collaboration with Miss Mackenzie, who found that using the correlations of weekly rainfall between different stations in Great Britain from a record of about 40 years, it is possible to estimate the correlation, for any one week of the year, with an accuracy which, for any single week, would have required records for over 1,000 years; this was due to the fact that the sequence of correlations could be well represented by a harmonic curve of only the second order, so that use could be made of some 2,000 actual simultaneous observations of weekly rainfall to eliminate errors of random sampling (8. Fisher and Mackenzie, 1922). In the present investigation the accuracy is limited by the number of crops recorded, but the method employed enables full use to be made of the meteorological data.

In the practical application of the method of the regression integral, it is not necessary that the rainfall record should be strictly continuous; indeed daily observations are themselves too numerous to be conveniently handled. The rain was therefore divided up into 61 periods of 6 days each, the integrations being replaced by summations over 61 terms. This introduces a slight modification into the form of the orthogonal functions of the time, for it is necessary for exact work that

$$\sum_1^{61} (T_r T_s) \qquad r \neq s$$

should vanish. The most convenient orthogonal functions to use are those developed by Esscher (3) in respect to mortality, and independently by the present author, in eliminating the slow changes observable in the wheat yields of the experimental plots

(5, 1921). If the time is measured from the middle point of a series of $n$ terms, in units equal to the interval between successive terms, we have

$$T_0 = \frac{1}{\sqrt{n}},$$

$$T_1 = \sqrt{\frac{12}{n(n^2-1)}}\, t,$$

$$T_2 = \sqrt{\frac{180}{n(n^2-1)(n^2-4)}}\,(t^2 - n_2),$$

$$T_3 = \sqrt{\frac{2{,}800}{n(n^2-1)(n^2-4)(n^2-9)}}\left(t^3 - \frac{n_4}{n_2}t\right),$$

$$T_4 = \sqrt{\frac{44{,}100}{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)}}\left(t^4 - \frac{n_6 - n_2 n_4}{n_4 - n_2^2}t^2 + \frac{n_2 n_6 - n_4^2}{n_4 - n_2^2}\right),$$

$$T_5 = \sqrt{\frac{698{,}544}{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)(n^2-25)}}\left(t^5 - \frac{n_2 n_8 - n_4 n_6}{n_2 n_6 - n_4^2}t^3 + \frac{n_4 n_8 - n_6^2}{n_2 n_6 - n_4^2}t\right),$$

where $n_2$, $n_4$, $n_6$ and $n_8$ stand for the mean values of $t^2$, $t^4$, $t^6$ and $t^8$. So that

$$n_2 = \tfrac{1}{12}(n^2-1),$$

$$\frac{n_4}{n_2} = \tfrac{1}{20}(3n^2-7),$$

$$\frac{n_6 - n_2 n_4}{n_4 - n_2^2} = \tfrac{1}{14}(3n^2-13),$$

$$\frac{n_2 n_6 - n_4^2}{n_4 - n_2^2} = \tfrac{3}{560}(n^2-1)(n^2-9),$$

$$\frac{n_2 n_8 - n_4 n_6}{n_2 n_6 - n_4^2} = \tfrac{5}{18}(n^2-7),$$

$$\frac{n_4 n_8 - n_6^2}{n_2 n_6 - n_4^2} = \tfrac{1}{1008}(15n^4 - 230n^2 + 407).$$

These functions are in reality very convenient to handle, the practical arithmetic involved will be explained in Section (5) under the analysis of the rain data. If $n$ be increased the functions tend to take the form of Legendre Polynomials.

### 4. *The Correlation of Residuals.*

When it is desired to study the correlation of variables such as annual figures, in which progressive changes are observable, it is necessary in order to obtain results which can be relied upon, and compared, in the same manner as those obtained from homogeneous material, to eliminate in some way the influence of the progressive changes. Material of this kind is very abundant and of the utmost importance ; the bulk of official statistics,

vital, economic, epidemiological, meteorological and agricultural, may be said to be awaiting a method of reduction which shall deal adequately with the difficulties which this type of data presents. My investigations have not led me to any satisfactory or complete solution of the problem, but since systematic methods have been developed and strongly advocated, it would seem worth while at the present time to put forward a group of mutually connected considerations, which shed new light upon various aspects of the problem, and bring clearly into view the sources of error to which the more obvious methods of approach are exposed.

It has been observed (4, FISHER, 1915) that if the individual values of a sample be taken as the co-ordinates of a point in generalised space, the mean standard deviation and coefficient of correlation of a sample are capable of a very beautiful geometrical interpretation. For if we assign any value $\bar{x}$ as the mean of $n$ observations $x_1, x_2, \ldots, x_n$, the equation

$$\mathrm{S}(x) = n\bar{x}$$

represents a plane section of the distribution, placed at right angles to the line

$$x_1 = x_2 = \ldots x_n$$

and meeting it at the point

$$x_1 = x_2 = \ldots = x_n = \bar{x}.$$

If this point is taken as origin, we have

$$\bar{x} = 0,$$

and the point representing any sample has rectangular co-ordinates

$$x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x},$$

so that its distance from the new origin is the square root of

$$n\mu^2 = \mathrm{S}(x - \bar{x})^2$$

where $\mu$ is the standard deviation of the sample.

The mean of the observations thus specifies the origin, and the standard deviation specifies the length of the radius vector from the origin to the sample point ; the correlational properties of the sample depend on the direction of this radius vector. For if a corresponding construction be made for the values of a second variate

$$y_1, y_2, \ldots, y_n,$$

and each of the axes of the $y$-space be brought into coincidence with the corresponding axes of the $x$-space, it has been shown, and it is easy to see, that the correlation coefficient between $x$ and $y$ is the cosine of the angle between the two *radii vectores.*

The geometrical interpretation of the correlation coefficient provides the simplest possible proof of a theorem of some importance in the present connection, namely, that : *If* $x_1', x_2', \ldots x_n'$ *be the co-ordinates of the first sample-point with respect to any rectilinear*

*orthogonal axes through the same origin and* $y_1'$, $y_2'$, ... $y_n'$ *the corresponding functions of* $y_1$, $y_2$ ... $y_n$ *then the correlation coefficient between* $x'$ *and* $y'$ *is equal to the correlation coefficient between* $x$ *and* $y$.

The system of orthogonal functions of the chief practical importance are those which arise in the elimination of progressive changes by means of a polynomial function of the time. These are the same functions as are used in analysing the sequence of weather in each particular season; it is worth noting that the above theorem is true of any orthogonal system.

In an unchanging series of independent values, *i.e.*, one in which the frequency of occurrence of any value is independent of the time, all the quantities $x'$, except $x_1'$, will be distributed in symmetrical curves about the mean value zero. If the values of $x$ are normally distributed about the mean $m$, with standard deviation $\sigma$, then all the values of $x'$ will be normally distributed with standard deviation $\sigma$, $x_1'$, about a mean $m\sqrt{n}$, and the remainder about zero. This follows at once from the fact that the distribution of the points representing the sample will be a globular normal cluster, and consequently the distribution of their projections on a line drawn in any direction will be invariable. Even when the values of $x$ are not normally distributed, those of $x'$ will in general be much more nearly so, owing to the properties of compound distributions which have often been discussed.

The simplest type of changing series is one in which the several values are distributed independently and in similar distributions, but about a changing mean. It is with a view to eliminating slow changes in the mean, that most methods of obtaining residuals from smoothed values are employed. It is usually the case that slow changes in the mean value may be sufficiently represented by a polynomial whose degree is small compared to the number of observations in the series. In such cases the first few values of $x'$ will be abnormally great (positive or negative), while the remainder will be normally and independently distributed. The whole group of values of $x'$ is heterogeneous, and the correlation of two such series will have no easy interpretation; its value may be governed by the values of a few exceptional terms, and therefore its probable error will be excessive. If, however, we reject the exceptional terms and calculate the correlation from the normally distributed remainder, we shall be correlating homogeneous material and the coefficient obtained will have a simple interpretation and its usual precision. Thus if the first $p$ values of $x'$ suffice to specify the slow changes in the mean value, the correlation required is that obtained by ignoring the first $p$ values of $x'$, and $y'$, or in other words putting them equal to zero.

This process is only an extension of the universal practice of eliminating the mean values of the variates in calculating the correlation coefficient; in working with an unchanging series we habitually ignore the values of $x'_1$ and $y'_1$. In the more general case we take

$$(n-p)\,\mu_1{}^2 = \sum_{p+1}^{n}(x'^2), \qquad (n-p)\,\mu_2{}^2 = \sum_{p+1}^{n}(y'^2), \qquad (n-p)\,\mu_1\mu_2 r = \sum_{p+1}^{n}(x'y').$$

From another point of view this may be regarded as the result of finding the partial correlation between $x$ and $y$ when the first $(p-1)$ powers of $t$, regarded as a system of correlated variates, are eliminated.

The value of $r$ obtained in this way is, as we have shown, identical with that obtained from the mean squares and mean products of the residuals of the two series after fitting polynomials of degree $(p-1)$. The values of such residuals are indeed heterogeneous in respect of variability, and mutually connected by $p$ equations; but since the values of $x'$ (and equally of $y'$) are independent and equally variable, the value of $r$ obtained from such residuals will be distributed in random samples, exactly as is that obtained from an ordinary correlation of $n-p+1$ independent homogeneous values. The frequency distribution of $r$ in such cases has been treated in detail (6, Fisher, '*Metron*,' 1921).

The elimination from data of obvious heterogeneity is, of course, a different process from a process which leaves the material necessarily homogeneous. The latter process is perhaps unobtainable, but it would be desirable, in cases where complications are expected, to have a means of testing whether the heterogeneity has been sufficiently eliminated. The elimination of the slowest changes is easy, since the earliest terms of the polynomial are quickly calculated. The size of the coefficients affords a criterion of rejection; thus for a record of the dressed grain from two plots on Broadbalk for 67 years I obtained (5, 1921).

| Dunged plot 2B. | | | Plot 11 (no potash). | | |
|---|---|---|---|---|---|
| Degree. | $x'/\mu$, | P. | Degree. | $x'/\mu$. | P. |
| 1 | −0·82 | 0·41 | 1 | −4·78 | 0·0000018 |
| 2 | −2·38 | 0·017 | 2 | +0·51 | 0·61 |
| 3 | −2·50 | 0·012 | 3 | +0·20 | 0·84 |
| 4 | −2·06 | 0·039 | 4 | −0·96 | 0·34 |
| 5 | +4·02 | 0·000058 | 5 | +1·72 | 0·086 |

This first column shows the degree of the polynomial term; the second, the value of $x'$ in terms of the standard deviation as estimated from the residuals after a curve of the 5th degree had been eliminated; the third column gives the probability of such a term occurring by chance.

Since the standard deviation is derived effectively from the sum of 61 independent squares, the exact form of the distribution of $x/\mu$, which is Student's Type VII curve, for $n = 62$, is taken to be equivalent to the normal distribution. On the dunged plot the first term gives no grounds for elimination; no significant deterioration having occurred on this plot, and it is immaterial whether or not such a term were retained in the correlations. The remaining four terms all show the existence of a slow change

in the yield. These four terms contribute to the variance over half as much as the remaining 61 terms put together. It should be noted that while little can be lost in rejecting the first term from the correlation, nothing can be gained by retaining it, for though its smallness renders it innocuous, its claim to rank as part of the homogeneous material is overturned by the size of the following terms. On plot 11 the deterioration due to exhaustion of potash is very great, and no other term except perhaps the 5th is large enough to excite suspicion. Our grounds for eliminating the first five terms in this case are that all the plots show similar and almost proportional slow changes, and that this is doubtless the case with Plot 11, where, however, the great variability of the annual yields has masked the significance of the coefficients.

Certainty as to the homogeneity of the remaining values of $x'$ could, I think, only be obtained by a complete calculation of their values; in the absence of tables of the coefficients this would be very laborious. The whole set should be normally distributed about zero, and heterogeneity would be indicated by $\beta_2$ significantly exceeding its normal value, 3. In the neighbourhood of the normal curve $\beta_2$ has been shown to be the most efficient statistic for such a test (7, 1922); a method of calculating $\beta_2$ for the $x'$ distribution from the actual values of $x$ would therefore be a valuable resource, especially if any way could be found of calculating it without the separate evaluation of the values of $x'$. It is believed that little heterogeneity remains in the series of wheat yields from Broadbalk after fitting polynomials up to the 5th degree.

The occurrence of series of values exhibiting slow changes is so widespread that it is not surprising that a number of different methods have been prepared for dealing with the difficulty, and that some difference of opinion should exist as to the disadvantages of each. We shall show that all the methods of elucidation which have been at all fully developed fall into one class, and that the main drawback of this class of methods is reduced to a minimum by the use of polynomials fitted to the whole series.

Slow changes may be eliminated in several ways :—

(i) by fitting a polynomial or other curve to the whole of the data ;
(ii) by the use of " smooth " values obtained by compounding a number of neighbouring terms ;
(iii) by repeatedly differencing the observed series.

In all these cases the resulting residuals designed for use as correlation variates may be successfully freed of any slow progressive trend which vitiates the original series; but in all cases this is achieved by entangling together to some extent the successive values. This effect is inevitable, for we must judge of the smooth value from which our residual is measured by the values at neighbouring epochs.

It might appear that simple differencing should be placed in a separate class, since here no smooth value objectively appears. There is no essential difference however; for if $2r$ is any even integer, we can construct a function,

$$v = u - (-)^r k\delta^{2r}u,$$

which is in effect the smooth value, the deviations from which are proportional to the $(2r)^{\text{th}}$ differences ; to make $\overline{v^2}$ a minimum for a series of equally variable quantities in random order we require

$$k = \frac{(2r!)^3}{(r!)^2\, 4r!}.$$

Thus the smooth value corresponding to 6th difference is

$$\begin{aligned} v_0 &= u_0 + {}_2\tfrac{5}{3}{}_1\, \delta^6 u_0 \\ &= {}_2\tfrac{1}{3}{}_1 \{131u_0 + 75\,(u_1 + u_{-1}) - 30\,(u_2 + u_{-2}) + 5\,(u_3 + u_{-3})\}. \end{aligned}$$

The formula so obtained is identical with that given by SHEPPARD (15, p. 31), for 7 point smoothing, using a polynomial of the 5th degree. The variate difference method is therefore only an extreme form of the use of SHEPPARD'S smoothing formula—the extreme in which the number of terms is a minimum for given degree of the slow change eliminated—while the other extreme is represented by the process of fitting a polynomial of the required degree to the whole of the series.

In treating these three processes as special forms of SHEPPARD'S smoothing process one distinction must be made. In fitting a polynomial to the whole of the series we wish to use not only the residual of the middle term, but the whole series of residuals. So in using smooth values from (say) a 15 point formula, the first and last seven residuals may be obtained from the curves fitted to the first and last sets of 15 points. In the applications which have been made of the Variate Difference Method, only the residuals of the middle terms have been used. The number in the series has been diminished by one with each differencing ; if however we wish to add the missing terms, this is easily done by means of binomial coefficients ; for example if

$$a,\ b,\ c$$

are the sixth differences of a series, the three missing residuals prior to $a$ are as follows

$$-\frac{a}{20},\quad +\frac{6}{20}a,\quad -\frac{15}{20}a,\quad a,\ b,\ c,\ \ldots\ .$$

The effects of such processes in entangling the neighbouring terms may best be seen by considering the effect of applying them to an unchanging series of equally variable quantities. If $u$ stand for such a quantity then, for example, the sixth difference may be written

$$v = -20u_0 + 15\,(u_1 + u_{-1}) - 6\,(u_2 + u_{-2}) + (u_3 + u_{-3}),$$

hence evidently

$$\begin{aligned} \overline{v^2} &= +924\,\overline{u^2}, \\ \overline{v_p v_{p+1}} &= -792\,\overline{u^2}, \\ \overline{v_p v_{p+2}} &= +495\,\overline{u^2}, \end{aligned}$$

and so on, the numerical coefficients being those of the expansion $(1+x)^{12}$. Consequently the correlations between neighbouring values of $v$ will be

$$\begin{aligned} r_1 &= -\tfrac{6}{7} &&= -0{\cdot}8571, \\ r_2 &= \tfrac{5}{8}\cdot\tfrac{6}{7} &&= +0{\cdot}5357, \\ r_3 &= -\tfrac{4}{9}\cdot\tfrac{5}{8}\cdot\tfrac{6}{7} &&= -0{\cdot}2381, \\ r_4 &= &&= +0{\cdot}0714. \\ r_5 &= &&= -0{\cdot}0130, \\ r_6 &= &&= +0{\cdot}0011. \end{aligned}$$

If such high correlations as these are produced in an originally uncorrelated series, it is clear that they cannot be used without drastic correction in an examination into such correlations between neighbouring terms as exist in the original series. Equally large are the effects when two separate series are correlated. If we take, for example, a second series $u'$, and obtain

$$v' = \delta^6 u',$$

then if the original two series were correlated so that

$$\overline{u_p u'_{p+k}} = \rho_k \sqrt{\overline{u^2}\,.\,\overline{u'^2}},$$

it is easy to see that the correlations obtained from the 6th differences will be expressible in terms of the $\rho$ series, in the form

$$r_k = \tfrac{1}{924}\,\delta^{12}\rho_k.$$

By such a method therefore we shall not obtain the true values $\rho_k$, but quantities proportional to the 12th difference of the series. Only if all values of $\rho$, except one, vanish, and we correlate the corresponding values of $u$ and $u'$, will we obtain an estimate of the correlation unaffected by gross inaccuracy. This constitutes a fatal objection to the applications which have been made of the variate difference method in its original form.

The same source of error still persists in more moderate degree when smoothing formulæ involving more terms are used. Let us take for example SHEPPARD's formula for fitting a polynomial of the 5th degree to sets of 15 points. The smoothed middle point is

$$\left(1 - \tfrac{50}{11}\,\delta^6 + \tfrac{75}{13}\,\delta^8 + \tfrac{36}{13}\,\delta^{10} + \tfrac{10}{17}\,\delta^{12} + \tfrac{15}{323}\,\delta^{14}\right) u\ ;$$

whence we have the residual

$$\begin{aligned} v = \frac{1}{11\,.\,13\,.\,17\,.\,19}\{&35126u_0 - 10125(u_1+u_{-1}) - 7500(u_2+u_{-2}) - 3755(u_3+u_{-3}) \\ &+ 165(u_4+u_{-4}) - 2937(u_5+u_{-5}) + 2860(u_6+u_{-6}) - 2145(u_7+u_{-7})\}. \end{aligned}$$

The correlations between neighbouring terms of an originally uncorrelated series now become

$$\begin{aligned} r_1 &= -0{\cdot}3075 & r_8 &= -0{\cdot}0132 \\ r_2 &= -0{\cdot}2370 & r_9 &= +0{\cdot}0072 \\ r_3 &= -0{\cdot}1119 & r_{10} &= +0{\cdot}0158 \\ r_4 &= +0{\cdot}0355 & r_{11} &= +0{\cdot}0099 \\ r_5 &= +0{\cdot}1440 & r_{12} &= -0{\cdot}0027 \\ r_6 &= +0{\cdot}1211 & r_{13} &= -0{\cdot}0076 \\ r_7 &= -0{\cdot}1565 & r_{14} &= +0{\cdot}0028. \end{aligned}$$

The correlations are now much more moderate. Their alterations are less violent, and since their sum is compelled to be $-0{\cdot}5$, their actual values are permitted to be considerably smaller. Such values are, however, sufficiently large to show that great inaccuracy will be introduced if we ascertain the mutual correlations of members of a series from those of the residuals from smooth values obtained from 15 adjacent points.

It is evident also that, apart from the question of the correlation of successive values, the correlations between two series will be vitiated in the same manner as by the variate difference method, though to a less degree; the correlations obtained may be expressed in terms of high differences of the true correlations, for if we write

$$\phi(\delta^2) = \tfrac{50}{11}\delta^6 + \tfrac{75}{13}\delta^8 + \tfrac{36}{13}\delta^{10} + \tfrac{10}{17}\delta^{12} + \tfrac{15}{323}\delta^{14},$$

and

$$\phi^2(\delta^2) \text{ for } \{\phi(\delta^2)\}^2,$$

then we have

$$r_k = \frac{\phi^2(\delta^2)\,\rho_k}{\phi^2(\delta^2)\,\rho_0}.$$

The general problem of eliminating the cross correlations from a pair of series showing slow changes is extremely complex; in the simplest case each value of the one series is correlated with only two adjacent values of the second series. To this case probably belongs the relation of wheat crop to weather, for the crop is admittedly much affected by the weather in the harvest year, and may to a much less extent be influenced, through the condition of the seed, or of the soil, by the weather of the previous year. It is fortunate, therefore, that one of the examples, upon which the largest amount of work has been expended, is probably also of the same type, and will serve to indicate the order of magnitude of the errors to be expected, through neglecting the cross correlation of the crop with the weather of the previous year.

The two series of infantile death rates in the first and second years of life, are probably connected principally, if not wholly, by the fact that the mortality, in any one year, of children in the second year of life, refers to nearly the same group of children as the mortality in the previous year of children in the first year of life; and by the second fact that mortality in the two age groups during the same year will be conditioned by the

same meteorological and epidemiological conditions. If the first effect is the main object of study, the latter will appear as a cross correlation introducing errors into our estimate of the first effect, according to the method of estimation employed.

In 1915 (1) ELDERTON and PEARSON found, using the variate difference method, the value $-0\cdot688$ for the correlation between the mortalities of the same group of children (males) in the two years. In 1923 the same authors using SHEPPARD's 15-point smoothing formula find the value $-0\cdot463$. The discrepancy is very great and suggests that the neglect of the correlation between the mortalities of the two groups of children in the same year has produced a large negative bias, which is more pronounced in the earlier estimate. Fitting a polynomial to the whole series further diminishes, though it does not eliminate, the error; for the polynomials of the 4th, 5th and 6th degrees we find the values $-0\cdot308$, $-0\cdot311$, $-0\cdot377$.

Now, assuming that only two correlations are really operative, these values may be corrected by calculating in a similar manner the apparent correlations in mortality for the two groups of children in the same year. These values are $+0\cdot4706$, $+0\cdot5456$ and $+0\cdot5217$; taking, then, the correlation between successive residuals of a series of $n$ terms fitted by a polynomial of degree $r$ to be $-(r+1)/(n-1)$, we have the equations

$$\text{4th degree} \begin{cases} \rho_0 - 0\cdot0877\rho_1 = -0\cdot3083 \\ -0\cdot0877\rho_0 + \rho_1 = +0\cdot4706 \end{cases} \qquad \rho_0 = -0\cdot2691;$$

$$\text{5th degree} \begin{cases} \rho_0 - 0\cdot1053\rho_1 = -0\cdot3109 \\ -0\cdot1053\rho_0 + \rho_1 = +0\cdot5456 \end{cases} \qquad \rho_0 = -0\cdot2563;$$

$$\text{6th degree} \begin{cases} \rho_0 - 0\cdot1228\rho_1 = -0\cdot3768 \\ -0\cdot1228\rho_0 + \rho_1 = +0\cdot5217 \end{cases} \qquad \rho_0 = -0\cdot3175.$$

The values of $\rho_0$ obtained from these equations are unbiassed estimates of the correlation required; they agree in indicating a correlation about $-0\cdot3$. This value may be confirmed from the figures given by PEARSON and ELDERTON for their 15-point smooth curve, which lead to the equations

$$\begin{aligned} \rho_0 - 0\cdot3075\rho_1 &= -0\cdot4548 \\ -0\cdot3075\,\rho_0 + \rho_1 &= +0\cdot6521 \end{aligned} \qquad \rho_0 = -0\cdot2808.$$

The concordance of these results indicates that whereas the variate difference method has exaggerated the value of this correlation to the extent of more than doubling it, the correlation of residuals from the 15-point smooth curve has reduced the error to about 50 per cent., and is moreover capable of correction provided only a few important correlations are present. The method of polynomial fitting has introduced errors of about 20 per cent. only, and the correction applied to it may be expected to be for this reason all the more precise.

In calculating the effects of weather upon the crop, it is probable, therefore, that the

effects of weather previous to the harvest year considered may be ignored. Hooker (10, 1907) has indeed indicated that some real effect may be ascribable to previous weather, but his correlations in the case of wheat are low, and as we have seen, if the polynomials fitted to the whole series are utilised to eliminate the slow changes, the spurious correlations introduced by this cause must be much further diminished. In the case of the infantile death rates when the correlation ignored was much greater than that which we sought to evaluate, the value obtained was raised only from $-0{\cdot}26$ to $-0{\cdot}31$ ; in the present case we may be certain that the ignored effects are much smaller than those evaluated, and their effect, if any, on the values obtained must be extremely small.

### 5. *The Rothamsted Rain and Wheat Data.*

The Rothamsted wheat data employed in this investigation are derived from Broadbalk field, which has grown wheat under experimental conditions since 1844. Thirteen plots have been continuously under uniform treatment since 1852. This series is unique in its length, which as we have seen is a most important consideration for statistical purposes. A second feature of great value is that the yields are derived from measured plots by actual weighing, and are not derived from estimates based on visual observation as are the county averages published by the Ministry of Agriculture. An account of the variations in yield of dressed grain of these 13 plots has already been published (5, Fisher, 1921). It was found that the variations observable could be divided into three groups, ascribable to three separate causes. (i) On many of the plots a progressive diminution is observable owing to the exhaustion of the soil in certain of the essential plant nutrients ; (ii) on all the plots slow changes in yield have taken place, which may probably be ascribed to variations in the weed infestation of the field as a whole ; this variation, unlike that ascribable to other causes, is in all the plots approximately proportional to the mean yield. These two causes of variation may be eliminated by fitting to each series a polynomial of the 5th degree. The remaining variation, consisting of the deviations of the actual yield from the smooth average value given by the polynomial, is ascribable primarily to variations of the season under which the crop grew to maturity. Conventionally this season has been taken to be the weather for 366 days, ending on August 31st of the harvest year. The first two causes of variation have already been discussed ; it is the purpose of the present paper to consider the fluctuations in yield from year to year in relation to the rainfall record.

Table II gives the manurial treatment of the 13 plots considered together with the mean yields obtained from them, and the mean rate of deterioration. In Table III are shown the deviations in each year from the corresponding polynomial value ; it will be observed that certain years have been omitted as not available for correlation with rainfall. The rainfall record commenced in February, 1853, and is therefore complete for the crop harvested in 1854 and subsequently. In 1889 and 1890 the field was partially fallowed by halves, and the crops in 1890 and 1891 following the fallow were

beneficially affected ; the same is true of the harvests of 1905, 1906 and also of 1915. It has been possible to use the value for 1916, since in this case the two halves of the field were harvested separately, and the half fallowed in 1915 could be rejected. With these omissions it is believed that the series of crop yields from this field are as homogeneous a series as it would be possible to obtain ; it will be seen that 60 years are available for correlation with rainfall.

TABLE II.

| Plot. | Manure per acre. | | | | | | Mean (Bushels per acre). | Mean annual diminution (Bushels per acre). |
|---|---|---|---|---|---|---|---|---|
| | Sulphate of potash. | Sulphate of soda. | Sulphate of magnesia. | Super-phosphate. | Sulphate of ammonia. | Chloride of ammonia. | | |
| | lb. | lb. | lb. | lb. | lb. | lb. | | |
| 2B | Dung | 14 tons | — | — | — | — | 31·549 | 0·031 |
| 3 & 4 | No Manure | | — | — | — | — | 12·269 | 0·097 |
| 5 | 200 | 100 | 100 | 392 | — | — | 14·180 | 0·090 |
| 6 | 200 | 100 | 100 | 392 | 100 | 100 | 22·581 | 0·141 |
| 7 | 200 | 100 | 100 | 392 | 200 | 200 | 31·367 | 0·144 |
| 8 | 200 | 100 | 100 | 392 | 300 | 300 | 35·694 | 0·092 |
| 10 | — | — | — | — | 200 | 200 | 19·504 | 0·157 |
| 11 | — | — | — | 392 | 200 | 200 | 22·046 | 0·219 |
| 12 | — | 366½ | — | 392 | 200 | 200 | 28·319 | 0·181 |
| 13 | 200 | — | — | 392 | 200 | 200 | 30·209 | 0·123 |
| 14 | — | — | 280 | 392 | 200 | 200 | 27·765 | 0·231 |
| 17 * | 200 | 100 | 100 | 392 | — | — | 14·510 | 0·092 |
| 18 * | — | — | — | — | 200 | 200 | 29·006 | 0·114 |

* Alternate.

In February, 1853, readings of rainfall were commenced with a large gauge, 0·001 acre, built for the purpose ; the readings of this large gauge have been consistently higher than those of the 5-in. and 8-in. gauges which are at present placed beside it. It may be concluded that the large gauge gives a better estimate of the amount of rain falling on the field. Daily readings are available for the whole period, the rain up to 9 a.m. being ascribed to the previous day. The rainfall was divided as explained in Section 3 into 61 periods of 6 days each, and each set of 61 values was then analysed by calculating the coefficients of the polynomials up to the 5th degree. Thus the amount and distribution of rain in each season was represented by a series of 6 numbers ; such a representation of a complicated sequence of events might seem to be insufficient, but it has been shown in Section 3, that only such coefficients are required as correspond to the regression function ; since the latter varies relatively slowly, little would be gained by following in more detail the rapid fluctuations of the weather.

The computation of the rainfall coefficients involved a great deal of labour. The method

## Table III.

| — | 2B. | 3 & 4. | 5. | 6. | 7. | 8. | 10. | 11 | 12. | 13. | 14. | 17 and 18. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | Min-erals. | Ammon. Sulphate. |
| 1854 | + 9·54 | + 6·86 | + 5·97 | + 7·62 | +11·95 | +13·98 | +12·79 | +15·83 | +13·16 | +12·98 | +12·44 | + 6·43 | +13·23 |
| | + 0·45 | + 1·33 | − 0·42 | − 0·06 | − 2·93 | − 5·65 | − 1·14 | − 7·67 | − 2·61 | − 2·80 | − 2·62 | − 0·67 | + 1·04 |
| 1856 | + 0·15 | − 1·59 | + 0·54 | − 1·34 | − 0·75 | − 0·03 | − 0·07 | + 0·91 | − 2·44 | − 3·45 | − 0·99 | − 1·66 | − 2·00 |
| | + 3·81 | + 3·89 | + 4·63 | + 5·69 | + 6·03 | + 7·82 | + 5·25 | + 8·10 | + 6·90 | + 6·98 | + 6·74 | + 6·26 | + 7·02 |
| 1858 | + 0·47 | + 1·14 | − 0·08 | − 1·05 | − 0·52 | + 0·51 | − 1·53 | + 0·63 | + 0·20 | + 0·38 | + 0·93 | + 1·39 | − 0·26 |
| | − 2·45 | + 1·45 | + 1·73 | + 0·01 | − 5·04 | − 7·39 | − 4·70 | − 4·10 | − 2·84 | − 2·62 | − 2·85 | + 0·08 | − 1·21 |
| 1860 | − 6·52 | − 3·44 | − 2·64 | − 7·61 | −11·87 | −10·85 | − 9·93 | − 8·96 | −10·19 | −10·56 | −10·08 | − 4·46 | − 8·53 |
| | − 3·79 | − 5·04 | − 2·74 | − 1·55 | − 4·26 | − 6·94 | −12·12 | − 6·74 | − 4·32 | − 2·89 | − 3·52 | − 1·20 | − 0·60 |
| 1862 | + 0·26 | − 0·06 | + 0·08 | − 0·32 | − 2·70 | − 2·20 | − 2·20 | − 3·74 | − 2·99 | − 4·21 | − 5·30 | − 0·92 | − 5·37 |
| | + 6·48 | + 2·86 | + 2·50 | +11·80 | +15·86 | +14·56 | +15·69 | +15·67 | +17·88 | +17·10 | +18·00 | + 2·37 | +13·66 |
| 1864 | + 3·20 | + 1·41 | + 0·21 | + 4·15 | + 8·88 | + 9·38 | + 9·05 | + 6·98 | + 9·23 | + 7·63 | + 6·28 | − 0·56 | + 4·10 |
| | + 1·12 | − 0·79 | − 1·94 | − 1·20 | + 4·29 | + 3·84 | + 3·36 | − 1·47 | + 0·07 | + 2·26 | + 2·12 | − 0·54 | − 0·09 |
| 1866 | − 2·53 | − 1·50 | − 2·34 | − 4·79 | + 5·01 | − 6·82 | + 3·57 | − 0·05 | − 5·31 | − 9·35 | − 5·63 | − 4·18 | − 4·54 |
| | − 6·82 | − 4·66 | − 5·82 | − 8·67 | −11·75 | − 7·72 | − 4·37 | − 5·23 | − 8·14 | − 9·51 | −10·00 | − 5·39 | − 6·79 |
| 1868 | + 8·25 | + 3·87 | + 3·01 | + 4·73 | + 6·92 | + 9·20 | + 3·78 | + 6·95 | + 8·13 | + 6·68 | + 9·78 | + 3·21 | + 7·58 |
| | + 5·51 | + 1·83 | + 1·61 | − 1·26 | − 3·58 | − 1·75 | − 2·08 | − 3·63 | − 3·58 | − 4·49 | − 3·40 | + 1·29 | − 6·38 |
| 1870 | + 4·45 | + 3·11 | + 5·00 | + 8·26 | + 9·41 | + 9·48 | + 1·32 | + 0·13 | + 5·29 | + 5·93 | + 5·38 | + 4·78 | + 5·46 |
| | + 7·55 | − 1·99 | − 1·44 | − 4·60 | − 8·10 | − 7·36 | −10·48 | −13·51 | − 8·08 | − 0·20 | − 5·34 | + 2·27 | − 0·02 |
| 1872 | + 1·43 | − 0·44 | − 0·20 | − 0·55 | + 0·01 | + 1·13 | − 1·59 | + 3·36 | + 0·87 | + 0·17 | + 1·38 | − 0·24 | − 2·18 |
| | − 3·80 | + 0·78 | − 0·01 | − 4·74 | − 7·18 | − 6·40 | + 0·72 | − 4·01 | − 4·78 | − 5·75 | − 4·25 | − 1·03 | − 7·50 |
| 1874 | + 8·97 | + 0·96 | + 0·52 | + 4·50 | +10·76 | + 7·04 | + 7·29 | +10·09 | +12·38 | + 8·38 | + 8·76 | + 1·54 | + 5·63 |
| | − 1·24 | − 2·59 | − 3·09 | − 3·74 | − 2·48 | − 3·21 | − 4·85 | − 4·38 | − 1·31 | − 1·16 | − 1·12 | − 0·47 | − 1·89 |
| 1876 | − 6·22 | − 2·03 | − 1·70 | − 4·22 | − 4·69 | − 3·30 | − 4·87 | − 7·59 | − 6·97 | − 2·85 | − 5·00 | − 1·55 | − 0·94 |
| | − 6·05 | − 1·17 | − 0·49 | − 5·34 | − 8·20 | − 8·04 | − 0·12 | − 3·71 | − 8·11 | − 9·54 | − 8·27 | − 1·86 | −14·80 |
| 1878 | − 2·13 | + 1·92 | + 2·55 | + 2·80 | + 3·15 | + 5·39 | +10·99 | + 8·27 | + 3·66 | + 1·81 | + 5·45 | + 3·47 | + 1·54 |
| | −14·69 | − 5·95 | − 6·53 | − 9·55 | −12·00 | −12·25 | −13·13 | − 9·94 | −11·34 | −11·58 | −10·29 | − 8·65 | − 7·26 |
| 1880 | + 7·27 | + 1·54 | + 5·43 | + 7·15 | + 5·97 | + 2·30 | − 5·25 | + 4·84 | + 4·03 | + 5·34 | + 4·50 | + 3·12 | + 4·99 |
| | − 1·35 | + 2·45 | + 0·30 | + 1·12 | − 2·17 | − 2·58 | + 1·76 | + 0·65 | − 1·56 | + 0·48 | + 0·91 | + 1·20 | + 3·93 |
| 1882 | + 0·58 | + 0·09 | − 0·07 | + 2·22 | + 6·54 | + 3·44 | + 7·69 | + 9·58 | + 9·28 | + 4·25 | + 8·10 | + 3·41 | + 2·69 |
| | + 2·44 | + 2·56 | + 2·96 | + 6·37 | + 6·36 | + 7·83 | + 0·71 | + 5·68 | + 5·24 | + 5·92 | + 6·37 | + 3·17 | − 0·46 |
| 1884 | − 1·00 | + 1·83 | + 2·52 | + 4·44 | + 8·16 | + 9·02 | + 8·80 | +11·35 | +10·08 | + 4·58 | + 9·26 | + 0·43 | + 4·80 |
| | + 5·90 | + 2·63 | + 2·01 | − 0·13 | + 0·74 | + 1·76 | + 7·17 | + 1·64 | + 1·46 | − 1·56 | − 0·23 | − 0·22 | + 3·57 |
| 1886 | + 1·53 | − 2·44 | − 2·15 | + 0·43 | + 3·84 | + 6·83 | − 4·49 | − 3·51 | + 0·28 | + 8·25 | + 3·68 | − 0·02 | + 7·95 |
| | − 0·98 | + 2·48 | + 0·85 | + 0·26 | − 2·47 | − 1·61 | + 4·24 | + 1·14 | + 3·59 | − 3·16 | + 1·27 | − 3·05 | + 0·48 |
| 1888 | + 1·54 | − 1·91 | − 1·93 | + 0·06 | + 3·32 | − 1·36 | − 5·87 | − 9·59 | − 3·14 | + 3·48 | − 1·43 | − 0·93 | + 1·33 |
| | + 3·33 | + 0·18 | + 0·99 | − 0·96 | − 2·46 | − 1·77 | − 5·71 | − 4·84 | − 2·95 | − 4·28 | − 3·26 | − 4·11 | − 7·69 |
| 1890 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 1892 | − 5·61 | − 3·41 | − 4·46 | − 2·84 | − 2·57 | − 0·74 | + 7·03 | − 5·42 | − 4·00 | − 2·56 | − 4·31 | − 2·95 | − 3·03 |
| | − 5·18 | − 2·28 | − 0·71 | − 5·79 | −14·55 | −17·45 | −10·31 | −13·48 | −16·97 | −15·29 | −15·64 | − 3·18 | −11·87 |
| 1894 | + 5·72 | + 4·80 | + 7·33 | +12·71 | +13·29 | + 9·41 | +11·30 | +17·87 | +18·61 | +15·72 | +15·77 | +11·57 | + 4·72 |
| | + 3·86 | − 2·50 | − 0·04 | − 4·22 | − 2·96 | + 0·09 | − 8·76 | − 5·19 | − 5·57 | − 3·49 | − 6·41 | − 3·26 | − 3·52 |
| 1896 | + 3·86 | + 3·39 | + 5·04 | + 4·55 | + 2·03 | + 4·15 | + 3·30 | + 3·05 | + 5·06 | − 4·58 | + 1·11 | + 1·10 | + 3·25 |
| | − 2·88 | − 3·74 | − 2·34 | − 5·81 | − 6·49 | − 3·20 | − 2·12 | − 4·84 | − 7·13 | − 4·72 | − 8·55 | − 4·98 | − 1·95 |
| 1898 | − 1·99 | − 0·43 | − 2·16 | − 4·21 | − 6·51 | −10·83 | + 1·65 | − 3·70 | − 4·12 | − 5·66 | − 3·50 | − 0·93 | − 6·27 |
| | + 2·78 | − 0·75 | − 2·18 | − 5·89 | − 3·28 | − 0·95 | + 3·81 | + 0·92 | − 0·10 | − 4·27 | + 0·88 | − 2·01 | − 5·29 |
| 1900 | − 6·06 | − 0·28 | − 0·04 | − 4·96 | − 4·30 | + 4·16 | + 0·77 | − 1·77 | − 4·00 | − 2·72 | − 3·66 | − 3·78 | − 2·75 |
| | + 0·85 | − 0·51 | − 0·21 | − 1·44 | − 3·95 | + 2·88 | + 1·76 | + 0·69 | − 1·77 | − 2·47 | − 3·87 | + 2·40 | − 2·86 |
| 1902 | + 3·41 | + 1·36 | + 1·05 | + 3·15 | + 5·50 | + 6·17 | + 5·00 | + 3·70 | + 5·20 | + 7·70 | + 6·67 | + 5·12 | + 5·85 |
| | − 7·60 | − 4·09 | − 5·47 | − 4·65 | − 5·29 | − 2·66 | − 4·98 | − 3·23 | −10·39 | − 4·71 | − 8·08 | − 8·75 | − 4·70 |
| 1904 | −14·09 | − 7·00 | − 6·83 | −10·16 | −11·51 | −12·87 | − 7·81 | − 5·98 | −11·42 | − 9·67 | − 7·14 | −10·67 | − 8·99 |
| | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 1906 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| | + 2·91 | − 0·82 | − 1·17 | + 4·78 | + 5·77 | − 0·34 | +10·45 | +15·10 | +16·01 | + 5·98 | +14·58 | − 1·98 | + 3·59 |
| 1908 | + 7·76 | + 2·90 | + 3·96 | + 3·76 | + 6·53 | +13·54 | + 5·04 | + 4·51 | + 7·95 | + 8·00 | + 4·93 | + 2·47 | + 6·73 |
| | + 2·90 | − 0·01 | − 1·48 | − 0·02 | + 3·21 | − 0·56 | − 5·89 | − 9·89 | − 4·63 | + 0·62 | − 4·36 | − 2·77 | + 4·39 |
| 1910 | − 0·70 | − 1·26 | − 1·46 | + 0·93 | + 0·87 | − 4·02 | − 1·71 | + 4·21 | + 2·80 | − 1·01 | + 3·97 | − 1·41 | + 1·76 |
| | + 7·92 | + 3·92 | + 3·65 | + 1·16 | + 1·76 | + 5·81 | + 7·31 | + 4·67 | + 4·17 | + 4·16 | + 4·86 | + 2·82 | + 3·61 |
| 1912 | − 9·77 | − 3·83 | − 5·25 | −13·03 | −16·12 | −18·99 | −12·68 | −12·33 | −17·11 | −18·62 | −15·95 | − 4·73 | −15·91 |
| | − 5·02 | − 3·10 | − 3·64 | − 3·89 | − 6·30 | − 3·26 | − 5·15 | − 3·77 | − 4·00 | − 5·79 | − 2·95 | − 3·37 | − 2·76 |
| 1914 | + 3·65 | − 2·51 | − 0·62 | + 1·91 | + 0·96 | + 0·23 | − 2·40 | − 0·49 | + 1·50 | − 4·16 | − 3·85 | − 5·11 | − 5·19 |
| | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 1916 | − 6·70 | − 3·70 | − 2·98 | − 4·60 | − 2·22 | −10·26 | + 4·11 | − 0·32 | − 3·55 | − 2·92 | − 3·81 | − 5·22 | − 0·94 |
| | −10·99 | − 1·48 | − 1·31 | + 2·50 | + 0·17 | + 4·03 | − 1·60 | − 3·42 | − 3·16 | + 4·39 | − 1·84 | + 2·82 | + 2·12 |
| 1918 | + 6·60 | − 0·03 | − 2·11 | − 3·19 | − 1·48 | + 2·05 | − 0·68 | − 0·93 | + 1·32 | − 0·74 | + 1·52 | + 2·11 | + 0·39 |

employed was that of successive summation, introduced by G. F. Hardy. By this method from a column of numbers $x_1, x_2, \ldots, x_n$ we obtain in succession

$$S_1 = x_1 + x_2 + x_3 + \ldots + x_n = S(x)$$
$$S_2 = nx_1 + (n-1)x_2 \ldots + 2x_{n-1} + x_n = S\{(n+1-r)x_r\}$$
$$S_3 = \frac{n(n+1)}{2}x_1 + \frac{(n-1)n}{2}x_2 + \ldots + 3x_{n-1} + x_n = S\left\{\frac{(n+1-r)(n+2-r)}{1.2}x_r\right\},$$

and so on. The successive summations carried as far as $S_6$ lead to very large numbers, but the work is straightforward and easily checked. It is not necessary to carry out the summation of the column in one piece. It may with advantage be broken in the middle; for example in the majority of the rainfall analyses the first 29 terms were summed as above, giving

$$z_1 = \overset{29}{\underset{1}{S}}(x_r)$$
$$z_2 = \overset{29}{\underset{1}{S}}\{(30-r)x_r\}$$
$$z_3 = \overset{29}{\underset{1}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on;}$$

while the remaining 32 values were summed backwards from the bottom of the column, dropping one term at the end of each summation, and so giving

$$z_1' = \overset{61}{\underset{30}{S}}(x_r)$$
$$z_2' = \overset{61}{\underset{31}{S}}\{(r-30)x_r\} = -\overset{61}{\underset{30}{S}}\{(30-r)x_r\}$$
$$z_3' = \overset{61}{\underset{32}{S}}\left\{\frac{(r-30)(r-31)}{1.2}x_r\right\} = \overset{61}{\underset{30}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on.}$$

Taking now alternately sums and differences, we obtain

$$S_1' = z_1 + z_1' = \overset{61}{\underset{1}{S}}(x_r)$$
$$S_2' = z_2 - z_2' = \overset{61}{\underset{1}{S}}\{(30-r)x_r\}$$
$$S_3' = z_3 + z_3' = \overset{61}{\underset{1}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on;}$$

from which the final sums of the whole column may be obtained from the equations

$$S_1 = S_1'$$
$$S_2 = S_2' + 32S_1'$$
$$S_3 = S_3' + 32S_2' + \frac{32.33}{1.2}S_1'.$$

The series $S_1, S_2, S_3, \ldots\ldots$, by whichever method it is obtained, is divided by the series of divisors

$$61, \quad \frac{61 \cdot 62}{1 \cdot 2}, \quad \frac{61 \cdot 62 \cdot 63}{1 \cdot 2 \cdot 3} \quad \text{and so on,}$$

yielding a series of numbers of similar magnitude

$$a = \frac{1}{61} \mathrm{S}(x_r)$$

$$b = \frac{1 \cdot 2}{61 \cdot 62} \mathrm{S}\{(n+1-r)\, x_r\}$$

$$c = \frac{1 \cdot 2 \cdot 3}{61 \cdot 62 \cdot 63} \mathrm{S}\left\{\frac{(n+1-r)\,(n+2-r)}{1 \cdot 2}\, x_r\right\}$$

with similar equations for $d$, $e$, and $f$. These may be regarded as weighted means of the series $x_1, \ldots, x_n$; it will be observed that the weights involve $r$, and therefore the time, up to the 5th degree. The series is therefore equivalent for the purpose of evaluating the polynomial coefficients to the first five moments of rainfall as distributed in time. The values at which we have arrived are, however, more convenient for computing purposes than the moments would be, for the quantities required by the method of Section 3 are

$$\rho_0 = \mathrm{S}(x\mathrm{T}_0) = \frac{1}{\sqrt{n}} \mathrm{S}(x) = \sqrt{n}\, a$$

$$\rho_1 = \mathrm{S}(x\mathrm{T}_1) = \sqrt{\frac{12}{n(n^2-1)}}\, \mathrm{S}(tx) = \sqrt{\frac{3n(n+1)}{n-1}}\, (a-b)$$

$$\rho_2 = \mathrm{S}(x\mathrm{T}_2) = \sqrt{\frac{180}{n(n^2-1)(n^2-4)}}\, \mathrm{S}\left\{\left(t^2 - \frac{n^2-1}{12}\right) x\right\}$$

$$= \sqrt{\frac{5n(n+1)(n+2)}{(n-1)(n-2)}}\, (a - 3b + 2c) \text{ and so on.}$$

The factors under the square root are not necessary for the correlational work, so that actually it is only necessary to calculate for each year

$$\left.\begin{aligned} a' &= a \\ b' &= a - b \\ c' &= a - 3b + 2c \\ d' &= a - 6b + 10c - 5d \\ e' &= a - 10b + 30c - 35d + 14e \\ f' &= a - 15b + 70c - 140d + 126e - 42f \end{aligned}\right\} \quad \ldots\ldots \quad \text{(V)}$$

These are the independent variates in terms of which the wheat yield is to be expressed

in the form of a regression equation. The regression coefficients of the yield upon these will have to be divided by factors of the form

$$\sqrt{\frac{(2s+1)\,.\,n(n+1)\ldots(n+s)}{(n-1)\ldots(n-s)}}$$

in order to give the coefficients $\alpha_s$ of Section 3, but as in the expansion of the regression function, $\alpha_s$ is multiplied by

$$\sqrt{\frac{(2s+1)\,.\,((2s)\,!\,)^2}{(s\,!)^4\,n(n^2-1)\ldots(n^2-s^2)}}$$

we actually multiply the regression coefficients by

$$\frac{(2s)\,!}{(s\,!)^2\,n(n+1)\ldots(n+s)}$$

in order to obtain the coefficients of the polynomial.

$$t^s+\ldots,$$

in the expansion of the regression function.

### 6. *The Rain Distribution Values.*

The values actually obtained for the quantities $a'$. ..., $f'$, for the 65 periods ending August 31st, 1854 to 1918, are given in Table IV; for tabulation they have been multiplied by 1,000; thus the value 347 in the second column shows that the average rainfall for the first period was 347 thousandths of an inch every 6 days. The figures in the third column measure the average rate at which rainfall was increasing or decreasing during the period, as indicated by a straight line fitted to the recorded values; $c'$ measures the parabolic term in the rainfall sequence, and so on taking more and more complex features of the distribution into account.

These values are themselves of very great interest, since the incidence of rainfall has not previously been analysed in this way.

The individual peculiarities of the successive seasons are brought out clearly, and it is possible to examine the sequence of years by adequate statistical methods. We may first enquire whether the observed sequence accords with the view that each season is an independent product of random causes under constant climatic conditions, or whether on the other hand the sequence indicates progressive changes in the quantity and distribution of the rainfall.

The series of 65 values of $a'$ was therefore analysed by summation in the same manner as the 61 values of the rainfall record of each year had been treated; as an illustration

TABLE IV.

| Harvest Year. | $a'$ | $b'$ | $c'$ | $d'$ | $e'$ | $f'$ | Harvest Year. | $a'$ | $b'$ | $c'$ | $d'$ | $e'$ | $f'$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1854 | 317 | − 14 | + 33 | − 16 | − 21 | − 1 | 1887 | 387 | − 82 | − 6 | + 29 | − 18 | +14 |
| 1855 | 398 | + 90 | + 41 | + 8 | − 35 | − 22 | 1888 | 500 | + 48 | + 59 | − 20 | − 10 | −13 |
| 1856 | 487 | − 18 | + 12 | + 1 | − 27 | + 35 | 1889 | 498 | + 72 | − 1 | − 11 | − 36 | −10 |
| 1857 | 404 | − 9 | + 19 | + 15 | − 9 | − 1 | 1890 | 450 | + 20 | + 35 | − 1 | − 15 | −12 |
| 1858 | 441 | − 69 | + 69 | − 38 | − 12 | + 15 | 1891 | 383 | + 65 | + 35 | + 22 | 0 | +28 |
| 1859 | 412 | + 64 | + 3 | − 14 | − 4 | − 5 | 1892 | 487 | − 29 | + 70 | + 45 | − 35 | +29 |
| 1860 | 598 | + 39 | + 35 | − 12 | − 12 | + 11 | 1893 | 398 | − 38 | + 37 | + 26 | − 16 | − 4 |
| 1861 | 381 | − 8 | + 3 | − 17 | − 24 | + 1 | 1894 | 488 | − 7 | + 24 | + 30 | − 8 | +16 |
| 1862 | 451 | + 22 | − 13 | − 13 | − 21 | + 18 | 1895 | 474 | − 3 | + 66 | + 51 | − 42 | −11 |
| 1863 | 414 | − 21 | + 36 | + 19 | − 7 | + 14 | 1896 | 399 | − 27 | + 21 | + 32 | − 4 | +13 |
| 1864 | 331 | − 56 | + 16 | − 24 | + 3 | + 3 | 1897 | 618 | −131 | + 77 | − 32 | + 55 | − 5 |
| 1865 | 454 | + 43 | + 61 | − 30 | + 29 | − 11 | 1898 | 323 | − 11 | + 9 | − 20 | + 4 | −17 |
| 1866 | 568 | − 2 | + 3 | + 38 | − 45 | + 8 | 1899 | 405 | − 20 | − 42 | + 27 | − 17 | + 5 |
| 1867 | 513 | − 14 | + 21 | − 27 | + 7 | − 17 | 1900 | 514 | − 47 | + 27 | + 35 | + 4 | + 2 |
| 1868 | 346 | − 11 | + 14 | + 26 | + 42 | − 11 | 1901 | 406 | − 10 | − 7 | + 31 | + 2 | +11 |
| 1869 | 426 | − 44 | − 37 | + 10 | − 3 | − 3 | 1902 | 382 | + 30 | + 18 | + 12 | − 5 | + 2 |
| 1870 | 351 | − 65 | + 10 | + 21 | + 9 | − 11 | 1903 | 520 | +106 | + 28 | − 12 | − 8 | −16 |
| 1871 | 451 | − 7 | + 16 | − 23 | − 32 | − 16 | 1904 | 517 | − 75 | + 30 | + 32 | + 2 | + 7 |
| 1872 | 475 | + 8 | + 13 | − 23 | + 5 | − 6 | 1905 | 428 | + 35 | + 58 | 0 | + 17 | +11 |
| 1873 | 503 | − 57 | + 6 | + 61 | − 31 | + 2 | 1906 | 389 | − 37 | − 13 | + 11 | − 9 | −15 |
| 1874 | 357 | − 12 | + 30 | − 10 | − 3 | − 6 | 1907 | 481 | − 44 | + 13 | + 5 | − 47 | +24 |
| 1875 | 526 | − 11 | + 42 | − 33 | − 26 | − 42 | 1908 | 494 | − 5 | + 6 | + 24 | − 2 | +38 |
| 1876 | 524 | − 80 | + 29 | + 18 | − 8 | + 12 | 1909 | 419 | + 52 | + 26 | − 26 | − 5 | − 7 |
| 1877 | 648 | − 76 | + 9 | + 22 | + 23 | − 38 | 1910 | 507 | − 16 | + 16 | + 23 | − 11 | + 5 |
| 1878 | 532 | + 38 | + 10 | + 20 | + 1 | + 35 | 1911 | 465 | − 43 | − 38 | + 25 | − 35 | +11 |
| 1879 | 674 | +105 | + 37 | + 26 | − 35 | − 3 | 1912 | 672 | + 18 | + 33 | +109 | − 9 | − 2 |
| 1880 | 350 | + 39 | + 25 | − 29 | − 7 | − 24 | 1913 | 451 | − 54 | − 35 | + 20 | − 2 | +18 |
| 1881 | 603 | − 78 | +110 | + 35 | + 41 | + 16 | 1914 | 419 | − 42 | − 14 | − 10 | + 8 | −57 |
| 1882 | 530 | − 11 | − 8 | + 5 | − 28 | + 15 | 1915 | 604 | − 11 | − 36 | + 65 | − 29 | −55 |
| 1883 | 573 | − 81 | + 6 | + 25 | − 39 | − 14 | 1916 | 582 | − 17 | − 26 | + 51 | + 15 | +14 |
| 1884 | 422 | − 53 | + 35 | + 1 | 0 | + 8 | 1917 | 577 | + 54 | + 82 | + 58 | − 25 | − 2 |
| 1885 | 431 | − 20 | − 37 | − 11 | + 9 | 0 | 1918 | 449 | − 18 | + 15 | − 22 | − 11 | +23 |
| 1886 | 508 | − 83 | + 33 | − 30 | − 13 | 0 | | | | | | | |

of the method the numerical values are given in Tables V and VI ; before summation each $a'$ was reduced by 400, the true value of the mean being inserted in the fourth column. The second column gives the successive sums, $S_1$ to $S_6$ ; from these the third column is derived by dividing by factors of the form

$$\frac{65 \,.\, 66 \ldots (65 + r)}{r!}, \qquad r = 1, \ldots, 5.$$

Thus the third column gives the values of $a$ to $f$ for the series $a'$. The fourth column gives the corresponding values $a'$ to $f'$ for the series $a'$, obtained by equations (V) ; while the fifth column, obtained by multiplying by factors of the form

$$\sqrt{\frac{(2r + 1) \,.\, 65 \ldots (65 + r)}{64 \ldots (65 - r)}}.$$

are the actual values of the first five transformed co-ordinates spoken of in Section 4, as $x_2'$ to $x_6'$.

TABLE V.—Analysis of sequence of values of $a'$.

| — | $S_1$ to $S_6$. | $a$ to $f$. | $a'$ to $f'$. | $x_2'$ to $x_6'$. |
|---|---|---|---|---|
| 1 | 4,521 | 69·55385 | 469·55385 | |
| 2 | 128,341 | 59·832631 | +9·72122 | +137·85 |
| 3 | 2,603,055 | 54·3378562 | −1·26834 | − 23·95 |
| 4 | 40,090,879 | 49·22841040 | +7·79456 | +182·35 |
| 5 | 500,901,859 | 44·570118750 | +2·35049 | + 66·31 |
| 6 | 5,313,933,165 | 40·528492120 | −2·62489 | − 88·43 |

In an unchanging series the values $x_2'$, $x_3'$ ... vary about zero in an approximately normal distribution, the standard deviation of which may be obtained from that of the original series ; for

$$\sum_2^n (x_r'^2)$$

is the sum of the squares of the deviations of the original series from their mean. Slow changes in the original series will be indicated by high positive or negative values in $x_2'$, $x_3'$, ... , and if such slow changes are suspected, it will be better to estimate the variance due to random causes from

$$\sum_7^n (x_r'^2),$$

from which the first five values have been omitted. From the sums of the squares of the deviations we may thus obtain a series of values each obtained from the last by deducting the square of the corresponding value $x_r'$; from each such sum may be obtained an estimate of the standard deviation due to random causes, by dividing by the number of squares concerned (degrees of freedom), and taking the square root. Such estimates will be equivalent to those derived from the residuals left after polynomials of the first to the fifth degree have been successively fitted ; but the labour of calculating the polynominal values is avoided.

TABLE VI.

| Degrees of freedom. | Sum of squares. | Mean square. | Standard deviation. |
|---|---|---|---|
| 64 | 465,105 | 7267·3 | 85·25 |
| 63 | 446,103 | 7081·0 | 84·15 |
| 62 | 445,529 | 7186·0 | 84·77 |
| 61 | 412,278 | 6758·7 | 82·21 |
| 60 | 407,881 | 6798·0 | 82·45 |
| 59 | 400,061 | 6780·7 | 82·35 |

The diminution of the estimates of the standard deviation indicates that the first 5 values of $x'$ are on the whole higher than those which follow; for example $x_4'$ is more than double the standard, and suggests strongly that real changes are taking place in the rainfall. To test this more accurately, divide the sum of the squares of the five deviations by the mean square of the remainder, then

$$\chi^2 = \frac{465105 - 400061}{6780{\cdot}7} = 9{\cdot}59,$$

whence, entering ELDERTON's table with $n' = 6$, we obtain P = ·089. Thus a larger value of $\chi^2$ would be obtained by chance only 8·9 times in a hundred, from a series of values in random order. There is thus some reason to suspect that the distribution of rainfall in successive years is not wholly fortuitous, but that some slowly changing cause is liable to affect in the same direction the rainfall of a number of consecutive years. Another way of putting the same result is that the variance estimated from the residuals of a polynomial of the 5th degree is 93·3 per cent. of the variance of the original series, so that some 6·7 per cent. of the total variance observed in annual rainfall may be ascribed to slow changes, while the remaining 93·3 per cent. of the variance are due to the chance circumstances of each particular year.

We summarise below the results of applying a similar analysis to the rainfall distribution values $a'$ to $f'$:—

TABLE VII.

| —— | $a'$ | $b'$ | $c'$ | $d'$ | $e'$ | $f'$ |
|---|---|---|---|---|---|---|
| Mean | 469·55 | − 11·09 | + 19·22 | + 9·43 | − 8·69 | + 0·57 |
| $x_2'$ | +137·85 | − 18·62 | − 35·01 | + 86·92 | + 8·78 | − 2·11 |
| $x_3'$ | − 23·95 | + 56·19 | − 33·75 | + 5·45 | − 16·94 | − 0·84 |
| $x_4'$ | +182·35 | − 35·31 | − 33·96 | + 13·02 | + 15·72 | − 23·80 |
| $x_5'$ | + 66·31 | + 13·00 | + 45·52 | + 5·76 | − 29·78 | + 1·25 |
| $x_6'$ | − 88·43 | + 24·69 | + 33·79 | − 20·47 | + 11·50 | + 22·33 |
| Standard Residue | 82·35 | 50·43 | 30·11 | 28·19 | 20·76 | 19·91 |
| $\chi^2$ | 9·59 | 2·17 | 7·42 | 10·33 | 3·78 | 2·70 |
| P | 0·089 | 0·83 | 0·19 | 0·067 | 0·58 | 0·74 |

It is quite clear, from the values of P. that no significant changes are observable in $b'$, $c'$, $e'$ and $f'$. These fluctuate with large standard deviations about the mean values

given in the first line of Table VII, which mean values have not significantly changed during the 65 years examined. With $a'$, and more clearly with $d'$, the mean appears to show gradual changes. In the latter case the change observable is very clear, and of a simple character. The value of $x_2'$, representing a linear increase, exceeds the standard deviation in the ratio 3·083; the probability of such a value occurring by chance is only 0·00205. The remaining values, $x_3'$ to $x_6'$, are individually and collectively insignificant, whence it appears that the slow change in $d'$ is significant, and may be represented by a uniform increase. This is in sharp contrast to the behaviour of $a'$, a contrast which is brought out by plotting successive 10 year means of these two quantities.

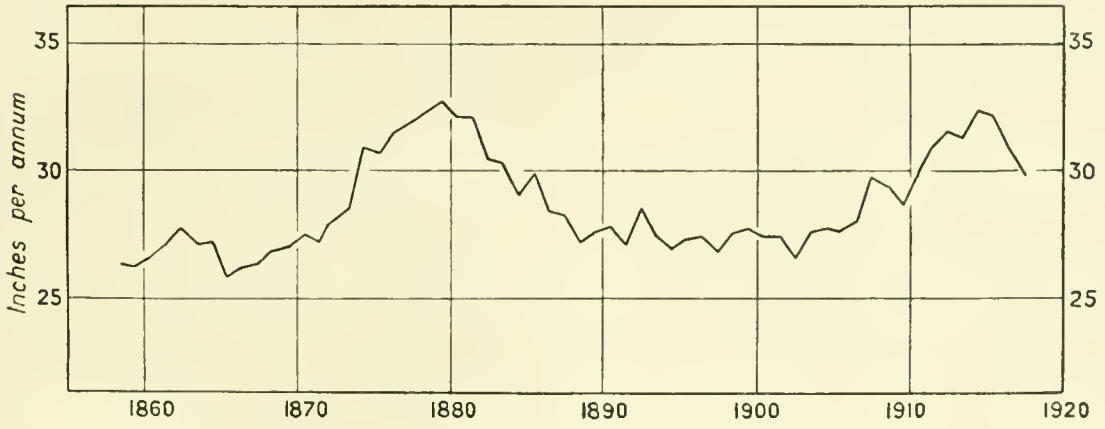
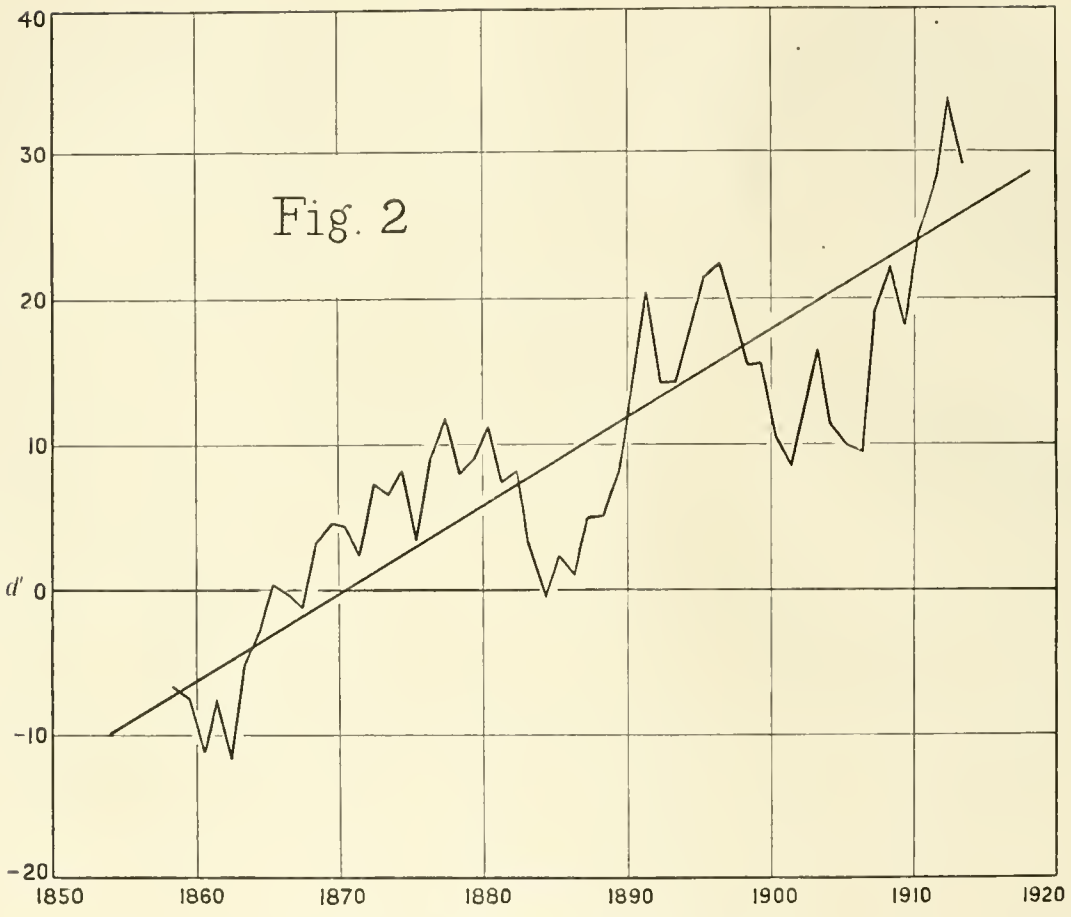



Fig. 1.

In the case of the total rainfall (fig. 1) measured by $a'$, there would seem to have been an excess of rain for a series of years about 1879, and a second spell of about equal intensity about 1914; the interval being 35 years. At these periods the annual rainfall average rises to about 32 inches. Prior to the first wet spell, and in the intermediate period, the average annual rainfall is about 27 inches. Such a change is small compared to the annual fluctuations; we have seen that it accounts for only about 6·7 per cent. of the variance, and that the whole effect is scarcely significant in our data. The distribution value $d'$, on the other hand, which measures not the total rainfall but its distribution over the year, shows a distinct and apparently uniform increase. Judging from the straight line (fig. 2), its mean value has changed from $-10$ in 1854, to nearly $+29$ in 1918; its mean value $+9·43$, given above, is thus not a permanent feature of the climate of the district. As in all rainfall features the annual fluctuation is very great: of the variation observed in $d'$ in the 65 years only 12·3 per cent. is ascribable to the linear change, the remaining 87·7 per cent. being apparently due to random fluctuations. In the absence of a similar analysis of rainfall at other stations, it would be premature to discuss the possible causes of this remarkable and progressive change in the climate. It may be remarked that such little additional information as is to be obtained from the monthly records indicates that the most marked feature of the change in progress is an




Ten year means of the distribution value $d'$.

increase in the December rainfall with perhaps some relative reduction of rain in Spring and Autumn.

The mean values of $c'$, $d'$, and $e'$, differ significantly from zero, owing to the seasonal variations of rainfall. The series of mean values, neglecting the mean value of $f'$ which is altogether negligible, represent the mean course of the rainfall sequence during the year. This is represented in fig. 3, where the mean monthly values have been inserted for comparison; these values have been reckoned *per day* to eliminate the effect of the varying lengths of the months. The monthly means are of course subject to large probable errors, and the smooth curve gives a slightly better estimate of the average sequence; this sequence is necessarily in process of modification owing to the progressive increase of $d'$.

The standard deviations of the six distribution values are in good mutual agreement and evidently arise from a common cause, namely the random fluctuation of rain in a six-day period. This may be seen by multiplying them by the factors necessary to transform $a'$, ..., $f'$ into $\rho_1, \rho_2, \ldots, \rho_6$ (Section 5), that is by

$$\sqrt{61}, \quad \sqrt{\frac{3.61.62}{60}}, \text{ etc.};$$

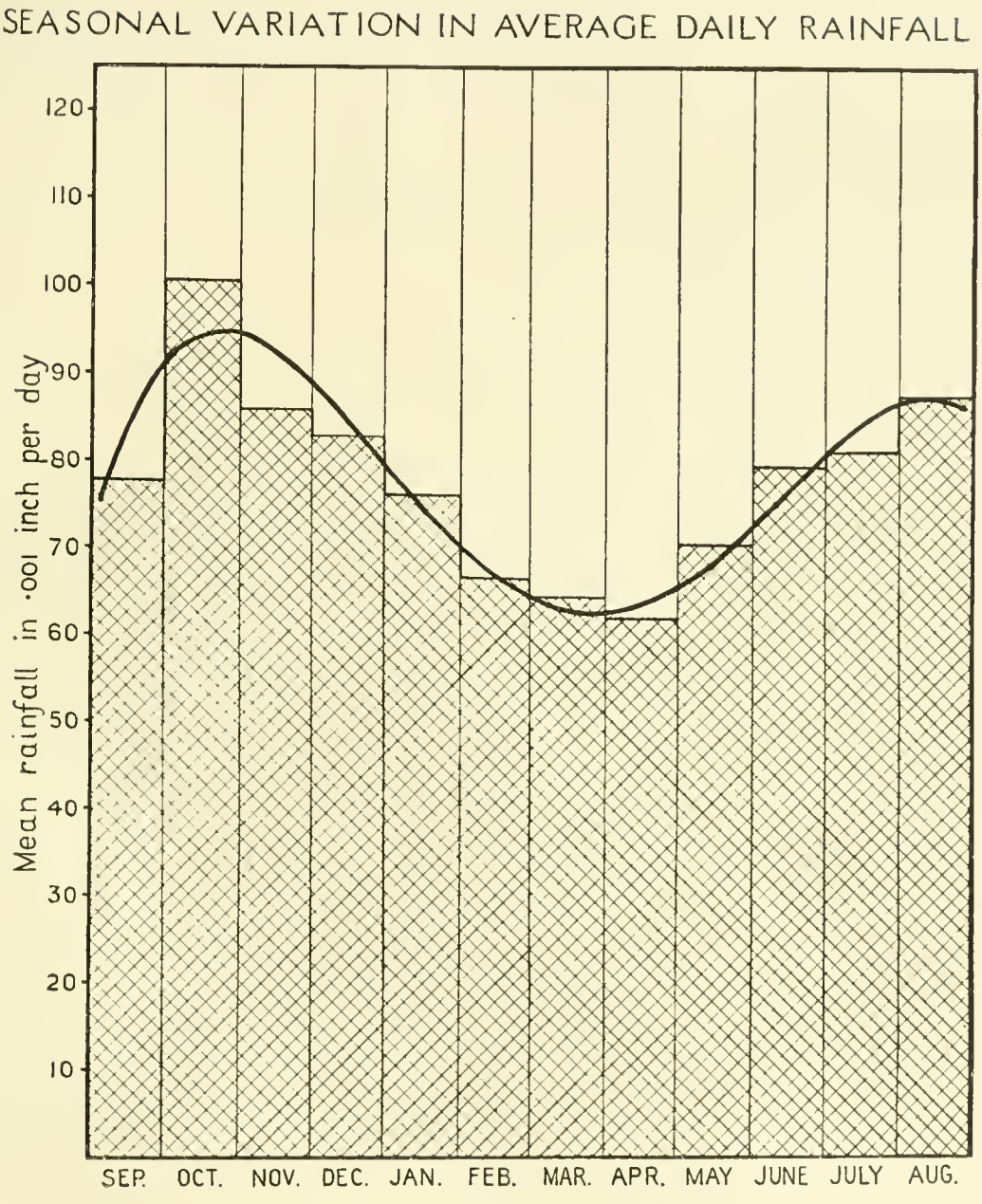



Fig. 3.

the standard errors of $\rho_1, \ldots, \rho_6$ are then

| $\sigma_{\rho_1}$. | $\sigma_{\rho_2}$. | $\sigma_{\rho_3}$. | $\sigma_{\rho_4}$. | $\sigma_{\rho_5}$. | $\sigma_{\rho_6}$. | Mean. |
|---|---|---|---|---|---|---|
| 643 | 693 | 552 | 643 | 573 | 660 | 627 |

These may be regarded as independent estimates of the standard deviation of the rainfall (in thousandths of an inch) in a six-day period, each derived from 59 independent squares, which is equivalent to a random sample of 60 values. The variation of these estimates is just of the order to be expected on the basis of random sampling. The

variation of the rainfall distribution values is thus nearly what we should deduce from the supposition that in addition to the small contributions made by secular and seasonal variation, each 6 day period varied independently with a standard deviation 0·627 inch, or a coefficient of variation of 134 per cent.

The distribution of the distribution values are not materially skew as far as can be judged from a series of 65 values. The following values obtained from the moments of the sample show the observed deviations from normality, with the standard errors appropriate to a normal population

| | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. | $f'$. | Standard error. |
|---|---|---|---|---|---|---|---|
| $\gamma_1 = \sqrt{\beta_1}$ | + 0·493 | + 0·336 | + 0·232 | + 0·566 | + 0·614 | − 0·546 | ± 0·306 |
| $\gamma_2 = \beta_2 - 3$ | − 0·290 | + 0·066 | + 0·354 | + 0·510 | + 0·920 | + 1·006 | ± 0·612 |

These figures suffice to show that the observed values hardly differ significantly from samples of normal distributions, although there are indications that, if the values for a considerably longer series of years were available, real departures from the normal distribution would become apparent.

### 7. *Correlations of Rainfall Distribution Values.*

The rainfall distribution values have been calculated from uncorrelated functions of the time, and if the rainfall at different periods of the year were uncorrelated the rainfall distribution values should also be uncorrelated. If, on the other hand, rainfall at one time of the year were associated with rainfall or lack of rainfall at another time, then correlations would appear in the rainfall distribution values, and such correlations would indicate the nature and extent of such mutual correlations between different parts of the season. Hooker observed, in the short period of 21 years at his disposal, a correlation as high as + 0·6 between the winter rainfall (1st to 8th weeks) and that of the previous autumn (37th to 44th weeks), and concluded that so large a correlation indicated a real interdependence between winter and autumn rainfall. More recently W. T. Russell (13, 1922) has calculated the correlation coefficients for monthly rainfall between consecutive and between alternate months. For Greenwich (65 years) no appreciable correlations are found between successive months, but in the alternate months it is found that June and August have a positive correlation + 0·55. Such a value should occur by chance from uncorrelated material once only in 900,000 times, so that even allowing for the fact that it is the highest of the 24 values obtained, it would appear to be significant of a real association. It should, however, be borne in mind that any secular changes either in total rainfall, or in its distribution through the year, as may exist at Greenwich, have not been eliminated in this coefficient, and, in the second place, the distribution of monthly rainfall is far from normal, and in consequence the probable error of such a determination may be distinctly higher than that calculated for normal distributions.

The correlations obtained between the 65 values $a', \ldots, f'$ are not, in any case, high. The transformed values ($z = \tanh^{-1} r$) are set out in Table VIII, since these values are distributed in random samples in approximately normal distributions, with the same standard error $\pm 0{\cdot}127$.

TABLE VIII.—Values of $z = \tanh^{-1} r$.

| —— | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. |
|---|---|---|---|---|---|
| $b'$ | −0·01 | — | — | — | — |
| $c'$ | +0·19 | +0·09 | — | — | — |
| $d'$ | +0·38 | −0·07 | −0·09 | — | — |
| $e'$ | −0·06 | −0·23 | +0·18 | −0·16 | — |
| $f'$ | −0·03 | −0·08 | +0·08 | +0·16 | −0·06 |

None of these correlations excite remark except that between $a'$ and $d'$; these two variates are the two which show secular variation, and it is therefore of more interest to obtain the corresponding values after the secular variation has been removed. These are found by treating the sum of the products of deviations as the sum of squares has been treated in Table VI, deducting successively the products of the values $x_2', \ldots, x_6'$, for the two variates concerned, calculating $r$ from the sums of squares similarly treated, and transforming to the $z$ scale as before. The values so obtained are shown in Table IX, with standard error $\pm 0{\cdot}1325$.

TABLE IX.—Values of $z$ after Eliminating Secular Change.

| —— | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. |
|---|---|---|---|---|---|
| $b'$ | +0·0342 | — | — | — | — |
| $c'$ | +0·2880 | +0·0800 | — | — | — |
| $d'$ | +0·3103 | −0·0526 | −0·0228 | — | — |
| $e'$ | −0·0784 | −0·2100 | +0·2380 | −0·1955 | — |
| $f'$ | +0·0356 | −0·1119 | +0·0381 | +0·2092 | −0·0525 |

The correlation of $a'$ with $d'$ now exceeds its standard error in the ratio 2·34; the probability of obtaining so large a value by chance from uncorrelated material is 0·019, and since this is the largest value of 15, it can scarcely be regarded as proof of association. To test the values collectively from the series of their squares, we obtain

$$\chi^2 = 22{\cdot}63, \qquad n' = 16, \qquad \mathrm{P} = 0{\cdot}093,$$

from ELDERTON'S Table, showing that although there are signs of association among the rainfall distribution values, such association, if it exist, is not strong enough to show up significantly in a series of about 60 values.

8. *The Regression of Yield on the Distribution Values, and on the Rainfall at Different Seasons.*

The sums of squares and products of the distribution values provide a basis for the calculation of the partial regression coefficients of the yields upon them, but before applying them in this way it is necessary to allow for the fact that 5 years' yields, namely, 1890, 1891, 1905, 1906 and 1915, have been omitted; we must therefore deduct from the sums the contributions of these years. Table X shows the polynomial values of $a'$, $b'$, ... , $f'$ for these years :—

Table X.

| Year. | $t$. | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. | $f'$. |
|---|---|---|---|---|---|---|---|
| 1890 | 4 | 465·64 | −13·21 | 33·17 | 9·11 | − 9·73 | 4·16 |
| 1891 | 5 | 461·66 | −12·36 | 33·36 | 9·10 | − 9·42 | 4·84 |
| 1905 | 19 | 458·63 | −11·96 | 10·57 | 18·88 | − 6·03 | 2·00 |
| 1906 | 20 | 463·62 | −12·52 | 7·98 | 20·29 | − 6·17 | 0·97 |
| 1915 | 29 | 524·83 | − 8·98 | 2·73 | 30·09 | −10·66 | − 4·13 |

calculated from the polynomial coefficients of Table VII. These values were subtracted from the observed values of Table IV, and the squares and products of the deviations deducted from the corresponding sums. The corrected sums, answering now to 54 degrees of freedom, were as follows :—

Table XI.

| — | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. | $f'$. |
|---|---|---|---|---|---|---|
| $a'$ | +380,853 | + 14,984 | +44,122 | +38,013 | − 5,278 | + 8,121 |
| $b'$ | + 14,984 | +140,110 | + 4,031 | − 4,241 | −14,577 | − 8,921 |
| $c'$ | + 44,122 | + 4,031 | +49,302 | + 906 | + 6,746 | − 1,392 |
| $d'$ | + 38,013 | − 4,241 | + 906 | +44,944 | − 5,790 | + 8,163 |
| $e'$ | − 5,278 | − 14,577 | + 6,746 | − 5,790 | +24,429 | − 2,769 |
| $f'$ | + 8,121 | − 8,921 | − 1,392 | + 8,163 | − 2,769 | +19,670 |

Since we require to find the partial regressions for 13 separate plots, it is worth while to invert the determinant formed of these numbers, so as to obtain a matrix of multipliers each of which is the co-factor of the corresponding number above, divided by the determinant. Table XII shows these multipliers in millionths.

Table XII.

| — | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. | $f'$. |
|---|---|---|---|---|---|---|
| $a$ | +3·235054 | −0·304445 | − 2·947149 | − 2·535496 | + 0·672526 | − 0·542793 |
| $b$ | −0·304445 | +8·094433 | − 1·082929 | + 1·042372 | + 5·775040 | + 4·100483 |
| $c$ | −2·947149 | −1·082929 | +24·070588 | + 0·730487 | − 7·637592 | + 1·050675 |
| $d$ | −2·535496 | +1·042372 | + 0·730487 | +26·747433 | + 5·214984 | − 8·794578 |
| $e$ | +0·672526 | +5·775040 | − 7·637592 | + 5·214984 | +48·604978 | + 6·478923 |
| $f$ | −0·542793 | +4·100483 | + 1·050675 | − 8·794578 | + 6·478923 | +57·557904 |

For each plot six correlation tables were constructed with the six rainfall distributions values, using the values in Tables III and IV, corresponding to the 60 years available. The six sums of products obtained from these tables, multiplied by the values of any column of Table XII and added, give the regression of yield on the corresponding rain variate. Table XIII gives the values obtained for these regression coefficients, the crop being measured in bushels per acre.

TABLE XIII.

| Plot. | Regression on— | | | | | |
|---|---|---|---|---|---|---|
| | $a'$. | $b'$. | $c'$. | $d'$. | $e'$. | $f'$. |
| 2B | − 40·9895 | + 2·1969 | − 8·0261 | − 46·2368 | − 15·8457 | + 24·0548 |
| 3+4 | − 20·2933 | − 0·0582 | + 0·1506 | + 13·7263 | + 14·2433 | + 27·0821 |
| 5 | − 20·7766 | − 3·1878 | − 0·0575 | + 13·9806 | + 7·2414 | + 18·8043 |
| 6 | − 39·7077 | − 0·6177 | + 34·1075 | − 5·7556 | − 8·3922 | + 51·2116 |
| 7 | − 42·4430 | − 3·1822 | + 36·3066 | − 31·9965 | − 16·6956 | + 83·5923 |
| 8 | − 43·8350 | − 11·7970 | + 35·9850 | − 31·6292 | + 8·1368 | +110·1433 |
| 10 | − 25·7070 | − 20·2730 | − 15·3172 | − 15·6733 | + 35·8060 | +140·2458 |
| 11 | − 30·4546 | − 14·8637 | − 9·8165 | + 1·6230 | + 17·7579 | +113·9780 |
| 12 | − 55·9573 | − 13·9676 | + 20·9149 | − 22·4509 | − 23·0060 | +113·0384 |
| 13 | − 45·1447 | − 3·9169 | + 63·5960 | − 48·0067 | − 24·4347 | + 92·2415 |
| 14 | − 42·6026 | − 12·9980 | + 25·9731 | − 30·7156 | − 24·8636 | +115·5578 |
| 17 & 18 miner'ls | − 24·0377 | + 0·5955 | + 17·2943 | + 5·6818 | − 3·8642 | + 33·4268 |
| 17 & 18 am'onia | − 42·6939 | + 3·8539 | + 60·9060 | − 44·3881 | + 2·7304 | +120·5257 |

These coefficients give directly linear regression formulæ expressing the deviation of the wheat crop on each plot in terms of the rainfall distribution values. Equally, as explained in Section 3, they enable us to estimate the average benefit or loss in bushels per acre ascribable to an additional inch of rain at any time during the year. For this purpose we divide the six regressions corresponding to any plot by

$$\frac{(r\,!)^2}{(2r)\,!}\,.\,61\,.\,62\,\ldots\,(61+r),$$

and we have the six coefficients of $t'$, $t$, $t^2$ ... in

$$a = \mathrm{A} + \mathrm{B}t + \mathrm{C}\,(t^2 - n_2) + \ldots,$$

expressing $a$ in terms of the polynomials of Section 3, wherein $t$ is the time in 6-day intervals measured from the central period of the year. Figs. 4 to 8 show the course of the function $a$ throughout the year.

### 9. *Discussion of Figs. 4 to 8.*

It should be emphasised that the information provided by a comparison of the rain record with the subsequent yields tells us the effect, not of so much rain, as such, but of

the total meteorological phenomena in fact associated with rain, at the time of the year considered. Thus in our records rain is associated with lower temperatures in summer, in winter with higher temperatures; and generally with diminished sunshine. The effects of these, to the extent in which they are associated, will be incorporated in the total effects shown in the charts, and are in fact an integral part of the value of a rain record as a means of foreseeing the prospects of the crop.

Even in the case of the rainfall itself, however, a detailed consideration of the ways in which it may affect the crop would lead to a most intricate discussion. It would be of the greatest value to know how important to the final crop, and how frequently influential, are not only the actual moisture available in the soil, but also the degree of soil aeration; how frequently, and to what extent, root development is hindered by soil saturation, or by an accumulation in toxic concentration of carbon dioxide. We should expect these factors to be intimately connected with rainfall, both in its direct effect in supplying fully aerated water, and in its indirect effects upon the soil texture.

In a single curve showing the average effect of rainfall (and of the average weather associated with such rainfall) upon the crop ultimately produced, all such contributory causes are included; by the comparison of the curves obtained from different plots, representing different manurial conditions of the soil, we may infer the effect of rainfall upon the availability of the manurial constituents of the soil. As will appear more fully, the predominant feature of such a comparison is the influence of excessive rain in removing soluble nitrates; this effect masks and overshadows all others, partly, I would suggest, by reason of its intrinsic importance to the wheat crop in our wet climate, partly because the plots on Broadbalk show great extremes in the relative abundance of available nitrogen.

The predominance in these curves of the effects which appear to be ascribable simply to the removal of soluble nitrate may perhaps explain two features which otherwise might be unexpected. In the first place the greater part of the effect of rainfall is expressible by means of the linear relations in the manure here represented; the quadratic terms, though of great interest and well worth further study, are of much less quantitative importance. Considering merely the effects of rain on the general environment of the growing plant we might expect definite optimum conditions throughout the year to be well marked, with correspondingly important quadratic terms; but the effect of nitrogenous fertilisers is not only approximately linear over the range concerned, but since increasingly heavy drainage generally removes decreasing quantities of nitrates, the tendency of its effect is to reverse the curvature of the regression. In the second place it might have been expected that the effects of rainfall would be closely related to the total crop on the different plots, and that resemblance would be more clearly apparent in the proportionate than in the absolute effects; but since for a wide range of manurial condition the absolute effects of additional nitrogen are not very unequal, and tend to decrease with increasing crop, it is apparent that the similarities in the absolute effect, which we observe for example in plots 8 and 14, are more probably due to quantitative

changes in nutrition than to indirect modifications of the physiological condition of the crop.

As will be seen from Table II, the 13 plots studied may be arranged in the following classes :—

(i) Low fertility, no added nitrogen. Plots (3 and 4), 5 and (17 and 18) mineral series.

(ii) Progressively higher yields due to progressive increase in nitrogenous dressings, from plot 6, through plots 7, 13 and (17 and 18) ammonium series, to plot 8.

(iii) Dressings which have produced an increasingly unbalanced nutrition with diminution of yield. Plots 12 and 14.

(iv) Unbalanced nutrition produced earlier and more intensely. Plots 10 and 11.

(v) Farmyard manure containing the very heavy contribution of about 200 lb. of nitrogen, with organic matter which has materially lightened the soil.

An examination of these diagrams shows how intimately the response of the crop to weather is connected with the manurial condition of the soil. Classing the plots solely by inspection of the curves of response to rainfall we shall put together every case in which the manurial treatment is alike, and indeed the whole series of curves arrange themselves in sequence of order of increasing abundance of nitrogenous fertilisers. As preliminary to a fuller discussion we may note (i) that in all the plots the average effect of additional rain is harmful. This agrees with HOOKER's finding for Eastern England. (ii) In all save the dunged plot 2B, the average loss is rapidly reduced during August, an observation which finds a simple explanation in the fact that the average date of carting the crop is August 24th, so that as August advances an increasing proportion of years occur in which the rain is too late to affect the crop. The exceptional behaviour of plot 2B indicates that the average loss per inch of rain in the month preceding harvest is even heavier than that shown in the diagram. (iii) In all the plots October is a month in which the average loss per inch of rain is small, or in which rain above the average is positively beneficial. This is the reverse of the condition found by HOOKER, who finds the greatest negative correlation with rain early in October. (iv) In all plots save two, the unmanured plot (3 and 4) and plot 5 which receives mineral (non-nitrogenous) manures only, the autumn period of benefit, or but little loss, from rain, is followed by a period centred in January in which dry conditions appear to be particularly desirable.

At this time of the year each additional inch of rain costs from one to two bushels in the crop. That this effect is scarcely visible on the unmanured plot, and still less so on plot 5, speaks strongly in favour of Sir JOHN LAWES's view that the damage done by winter rain was principally occasioned by the washing out of nitrates from the soil. Plot (3 and 4) and especially plot 5 can have little to lose in this way, and possibly rely for the greater part of their growth upon nitrates produced by bacterial activity as the soil grows warmer in March, April and May. The same view is confirmed by the fact that among the remaining plots the loss due to winter rain is least in plots 2B, 10 and 11, all of which are characterised by the low proportion utilised of the nitrogen supplied.

If these plots seldom suffer from lack of nitrogen, its loss during the winter must be expected to affect the crop relatively little; nevertheless the actual average loss of about one bushel to the inch of rain shows that at this time (*i.e.*, before the spring dressings are applied to plots 10 and 11, and perhaps before bacterial activity can fully tap the resources of the farmyard manure), a temporary shortage sometimes injures the crop. With these preliminary observations we may pass to the consideration of the types of rainfall response curve actually found.

Type I.—Plot 5, mineral manures only, and plot (3 and 4), unmanured. Fig. 4 shows plot 5.

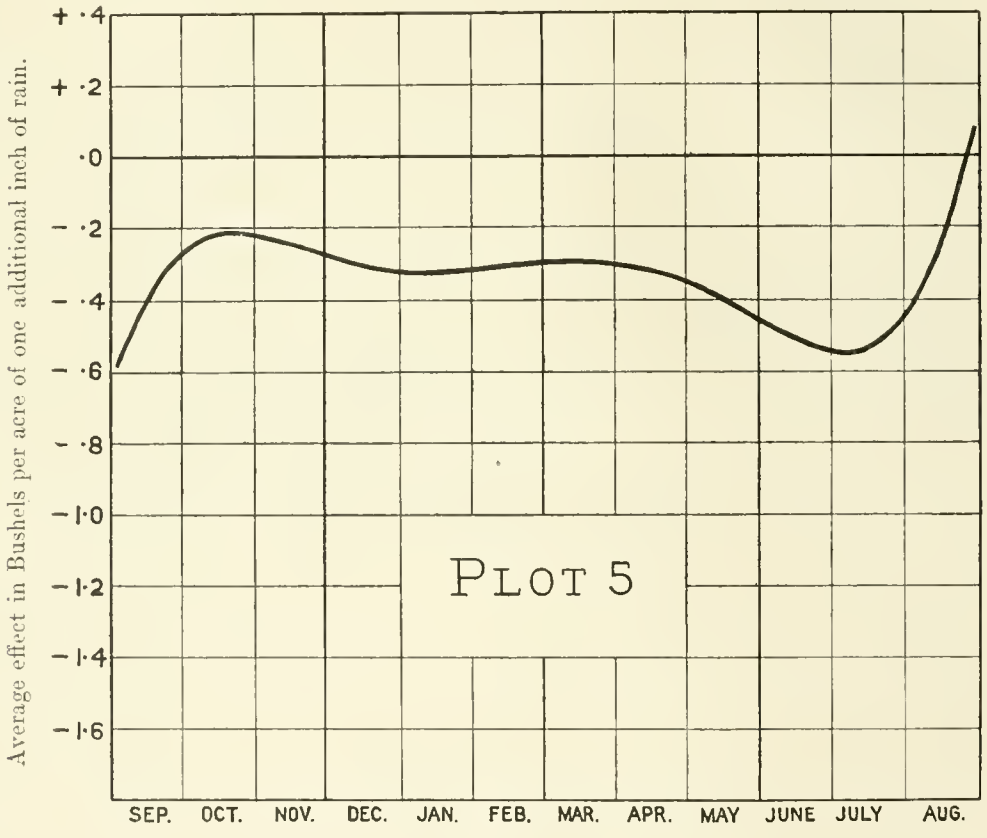



Fig. 4.

In these two plots the crop is severely limited by the lack of available nitrogen. Plot 5 has a mean yield of 14·18 bushels per acre, while plot (3 and 4) yields 12·27. These are averages (1852–1918) covering the whole period over which the rainfall regressions have been calculated. Both plots show heavy deterioration over this period. The small difference in the mean yields, which is nearly constant over the whole period, shows that lack of nitrogen is the dominant limitation in plot (3 and 4) as in plot 5 where the limitation due to lack of nitrogen must be most stringent.

In plot 5 the rain experienced is in excess of the optimum probably at all times of the year. The average loss for additional rain falls from about 0·4, in the September before sowing, to a minimum of 0·22 (1·5 per cent.) towards the end of October, and

remains steadily near to 0·30 till the end of May. It falls to a maximum loss of 0·55 (3·9 per cent.) early in July and rapidly diminishes through August. Since the proportion of drainage through the twenty-inch gauge decreases from nearly 80 per cent. in the winter to about 25 per cent. from April onwards, it follows that if the main loss on this plot due to rainfall is ascribed to the washing out of nitrates, the loss due to an equal amount of drainage is increasingly serious as the year advances. This accords with the belief that nitrification becomes more active in the Spring, and that this plot relies largely for its nitrogen supply upon bacterial activity at this season. The curve even suggests the possibility that under the intense nitrogen starvation of this plot the supply of nitrates is of importance as late as the beginning of July, though it must be borne in mind that other effects of rain may become important during the summer.

The curve for plot (3 and 4) is closely similar. September damage of about 0·4 is followed by a minimum of 0·23 (1·9 per cent.) in October. There is slight evidence of a maximum of damage in the winter of 0·37 (3·0 per cent.) followed by a minimum damage of 0·24 (2·0 per cent.) at the end of March ; this feature becomes marked in plots in which the nitrogen limitation is less acute, and is perhaps an indication that this condition is less extreme in the unmanured plot than in plot 5. The maximum damage in early summer is 0·58 (4·7 per cent.) ; little importance can be attached to the positive values of the last 12 days of August, save as in indication that August rain is inoperative after the first week of the month.

Type II.—Plot 6, complete minerals with single dressing of ammonium salts applied half in Autumn and half in Spring ; plot 17 and 18 mineral series, receiving mineral manures only, but alternating with the ammonium series which receives a double dressing of ammonium salts. Fig. 5 shows plot 6.

The similarity of the curves for these two plots is striking, and requires that the interpretation put on the yields of the alternating plot should be reconsidered. Plot 6 has a mean yield of 22·58 bushels to the acre, and shows a relatively rapid deterioration ; 17 and 18 minerals has a mean yield 14·51, scarcely more than that of plot 5 which receives the same dressing ; its average deterioration also resembles plot 5. Hence it has been thought that no appreciable benefit accrued to the alternating series from the previous year's dressing of ammonium salts. It has however been observed (5, 1921) that the alternating series is much more variable in yield than is plot 5, and this suggests that the additional variation is due to variable residue of nitrogenous material. This suggestion is strongly confirmed by the resemblance to plot 6, especially in the effects of winter rainfall.

Plot 6 has a strongly marked minimum of damage in early October, when the loss is only 0·10 bushel per inch of rain, or 0·4 per cent. ; this is followed by a winter period in which rain is on the average particularly harmful to the extent of 1·14 (5·0 per cent.) ; from April to July the average damage is nearly constant at about 0·64 (2·8 per cent.).

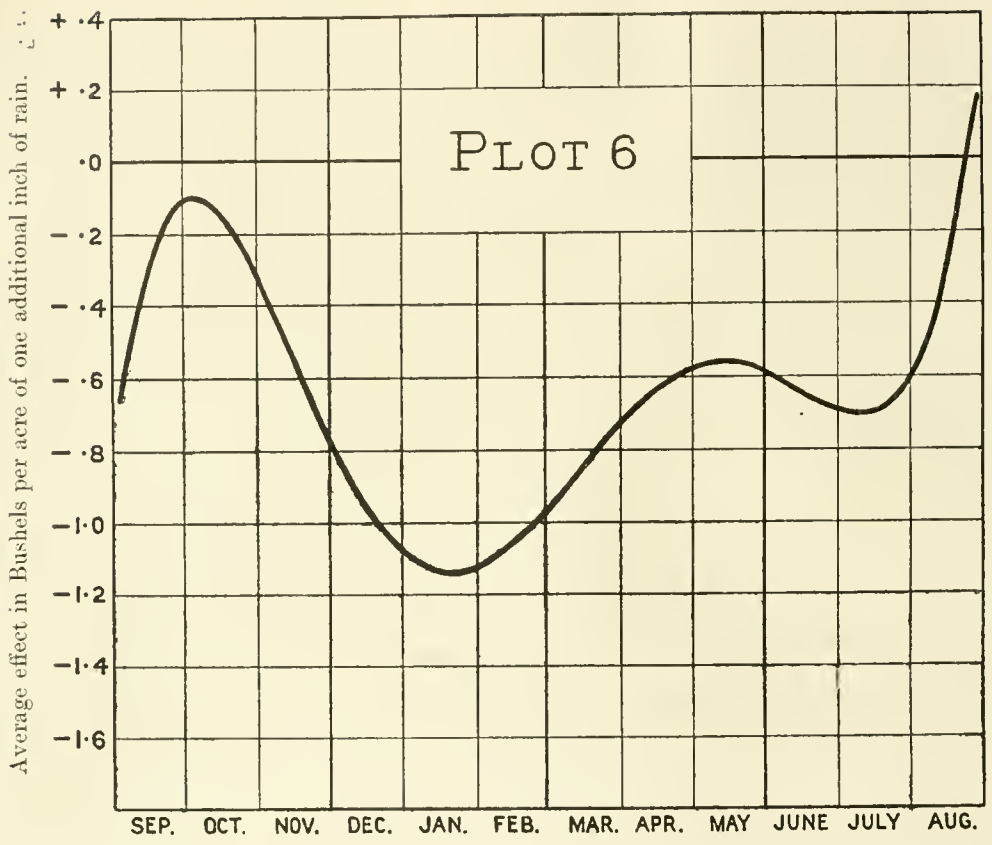



Fig. 5.

The effects on 17 and 18 minerals are, on the whole, smaller, being approximately proportional to the lower yield. A minimum of damage occurs in October 0·10 (0·7 per cent.), followed by a winter maximum 0·64 (4·4 per cent.), and by a nearly stationary period from April to July at 0·42 (2·9 per cent.). The winter effect is smaller both absolutely and relatively than that of plot 6, but if it be admitted that the importance of winter rain lies largely in the washing out of nitrates, the evidence for considerable, though no doubt variable, residues in the alternating series is unmistakable. The fact that the alternating series has not given an appreciable higher yield than that of plot 5, must be ascribed to soil heterogeneity, which has approximately balanced the advantage of the residual nitrogen. That such soil heterogeneity exists in this field may be shown by a comparison of 2*a* with the adjacent plot 2*b*. These have received the same manure since 1885, but during the present century 2*a* has ceased to gain in yield upon 2*b*, but yields very regularly on the average 2 bushels less than the adjacent plot. In view of this fact we cannot deny the possibility that with equal manuring the land of plots 17 and 18 would yield 2 or more bushels less than that of plot 5, and that this circumstance has served to mask the advantage accruing to the mineral series owing to residual nitrogen. It may be mentioned in addition that the mean yield of 17 and 18 ammonium series is 2·36 less than that of plot 7, a difference which may be ascribed to the combined effect

of soil inferiority and lack of residual nitrogen ; it should be borne in mind that a difference of this amount in a yield of 30 bushels probably would require the addition of considerably more nitrogen than an equal increase in a yield of 14 bushels.

Type III.—Plot 7, complete minerals, with double dressing of ammonium salts, of which one quarter is applied in autumn.

Plot 13, as plot 7 without sulphates of sodium and magnesium.

Plot 17 and 18 ammonium series.

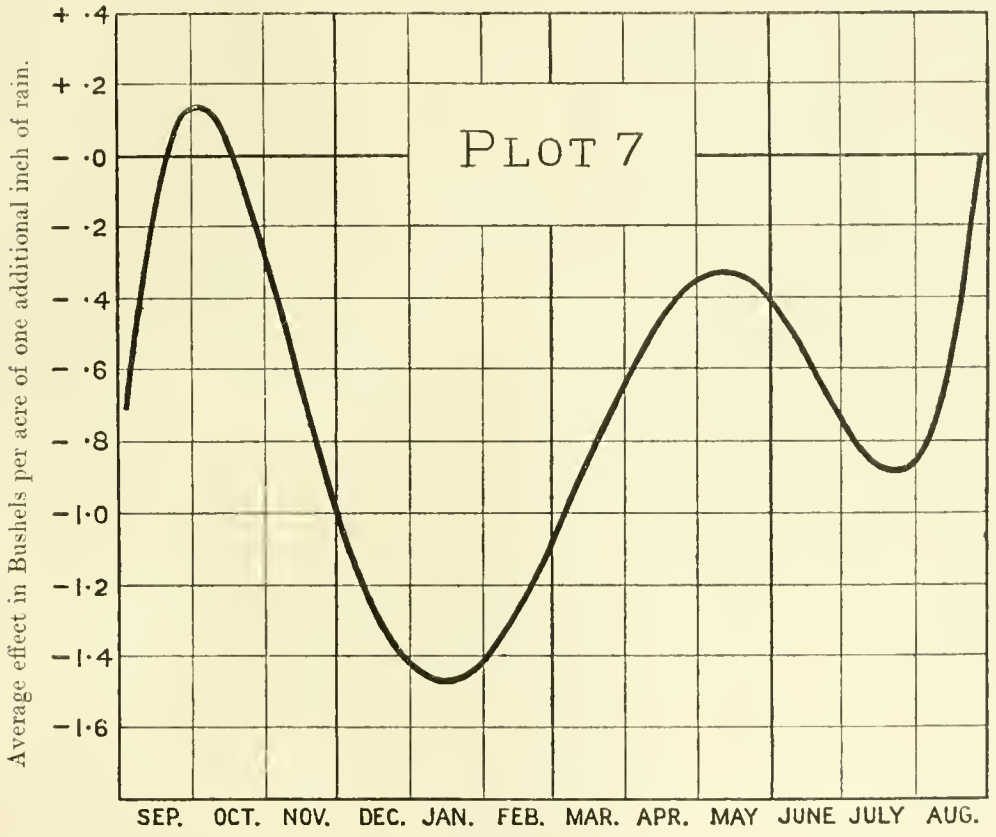



Fig. 6.

These three plots receive similar manurial treatment ; the mean yields are 31·37, 30·21 and 29·01 bushels per acre, with mean annual decrements 0·144, 0·123 and 0·114. It might have been anticipated that the decrement should be least on the alternating plot ; for the rest it is not obvious to what circumstances the small difference between the plots are to be ascribed.

The three rainfall curves are generally similar, all indicate a slight benefit for something over a month in autumn, the maxima being + 0·14 (0·4 per cent.), + 0·43 (1·4 per cent.), and + 0·41 (1·4 per cent.) ; the winter damage is strongly marked with values 1·47 (4·7 per cent.), 1·82 (6·0 per cent.) and 1·85 (6·4 per cent.). There is in all cases a clear period of minimum damage in May, the values falling to 0·34 (1·1 per cent.), 0·34 (1·1 per cent.) and 0·20 (0·7 per cent.), followed by a second period of maximum damage in July which is however less severe than the winter maximum, the values are

0·89 (2·8 per cent.), 0·79 (2·6 per cent.) and 0·94 (3·2 per cent.). On the average plot 13 suffers somewhat more heavily from rain than do plots 7 and the ammonium series, but in considering the effects at different times of the year it is seen to agree with the latter closely during the first half of the year, and more nearly with plot 7 during the second half.

Type IV.—Plot 8 as plot 7 with additional ammonium salts applied in the spring.

Plot 12 as plot 13 with substitution of sodium for potassium sulphate.

Plot 14 as plot 13 with substitution of magnesium for potassium sulphate.

Fig. 7 shows plot 8.

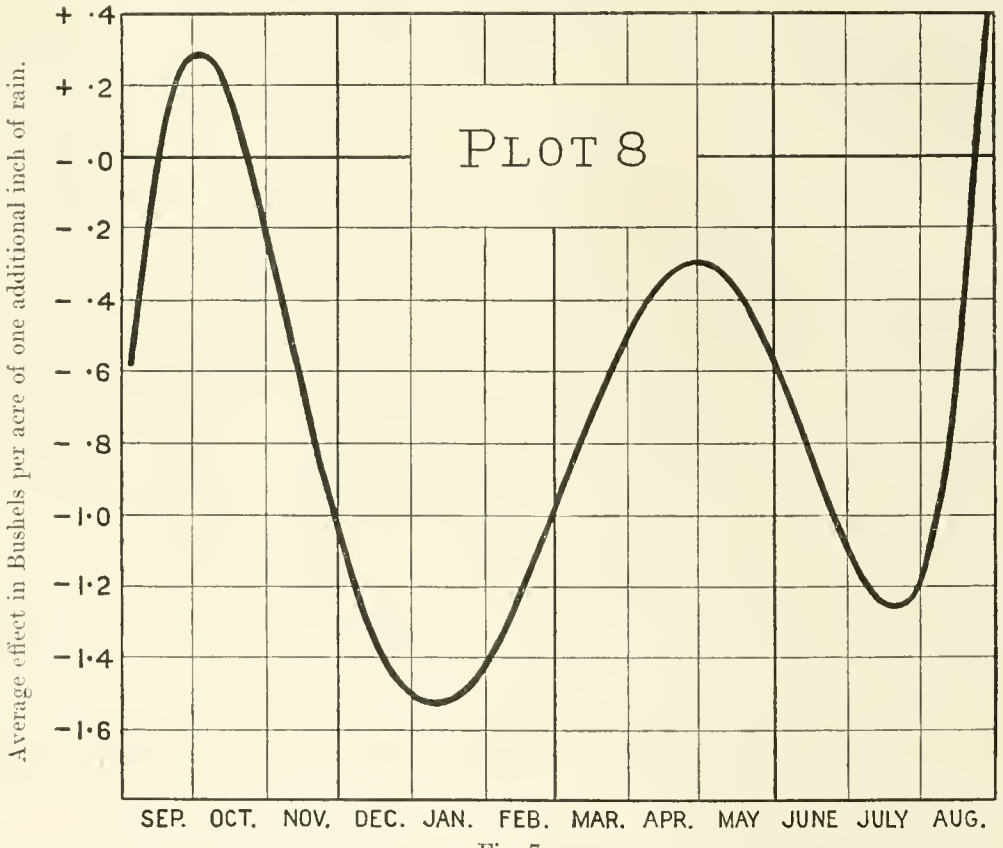



Fig. 7.

In most respects plot 8 contrasts strongly with plots 12 and 14, which however it resembles in its response to rainfall. The mean yield is 35·69 against 28·32 and 27·77, while it shows but slight deterioration, 0·092 annually, as against the heavy deterioration. 0·181 of plot 12, and 0·231 of plot 14. The resemblance consists in the greater relative abundance of nitrogen supplied to these plots than to those in Type III, and the correspondingly less relative abundance of nitrogen than those in Type V. In the case of plot 8, although the yield is greater than that of plot 7, yet since it receives a triple in place of a double dressing of ammonium salts, it is clear that this plot must less frequently suffer from lack of nitrates. Plots 12 and 14 on the other hand receive the same nitrogenous dressing as plot 7, but owing to lack of potash yield some 3 bushels an acre less ;

consequently with them also crop limitation through lack of nitrates must be less frequent. The same argument applies still more strongly to plots 10 and 11 in the next type.

The benefit of October rain is relatively large in plots 8 and 14, and plot 12 also reaches positive values at this period, the three values are + 0·29 (0·8 per cent.), +0·10 (0·3 per cent.) and + 0·36 (1·3 per cent.); the winter damage is strongly marked, though on the whole less so than in the preceding group, 1·53 (4·3 per cent.), 1·65 (5·8 per cent.), 1·52 (5·5 per cent.). The chief contrast lies in the deepening of the second period of maximum damage; in consequence of this the time of minimum summer damage is shifted back to the beginning of May with values 0·30 (0·8 per cent.), 0·49 (1·7 per cent.) and 0·23 (0·8 per cent.); while the second maximum reaches in late July the values 1·27 (3·5 per cent.), 1·48 (5·2 per cent.), 1·23 (4·4 per cent.), showing that dry weather at this period is to these plots nearly as beneficial as in January.

Type V.—Plot 10 double dressing ammonium salts only.

Plot 11 double dressing ammonium salts with superphosphate. Fig. 8 shows plots 10 and 2B.

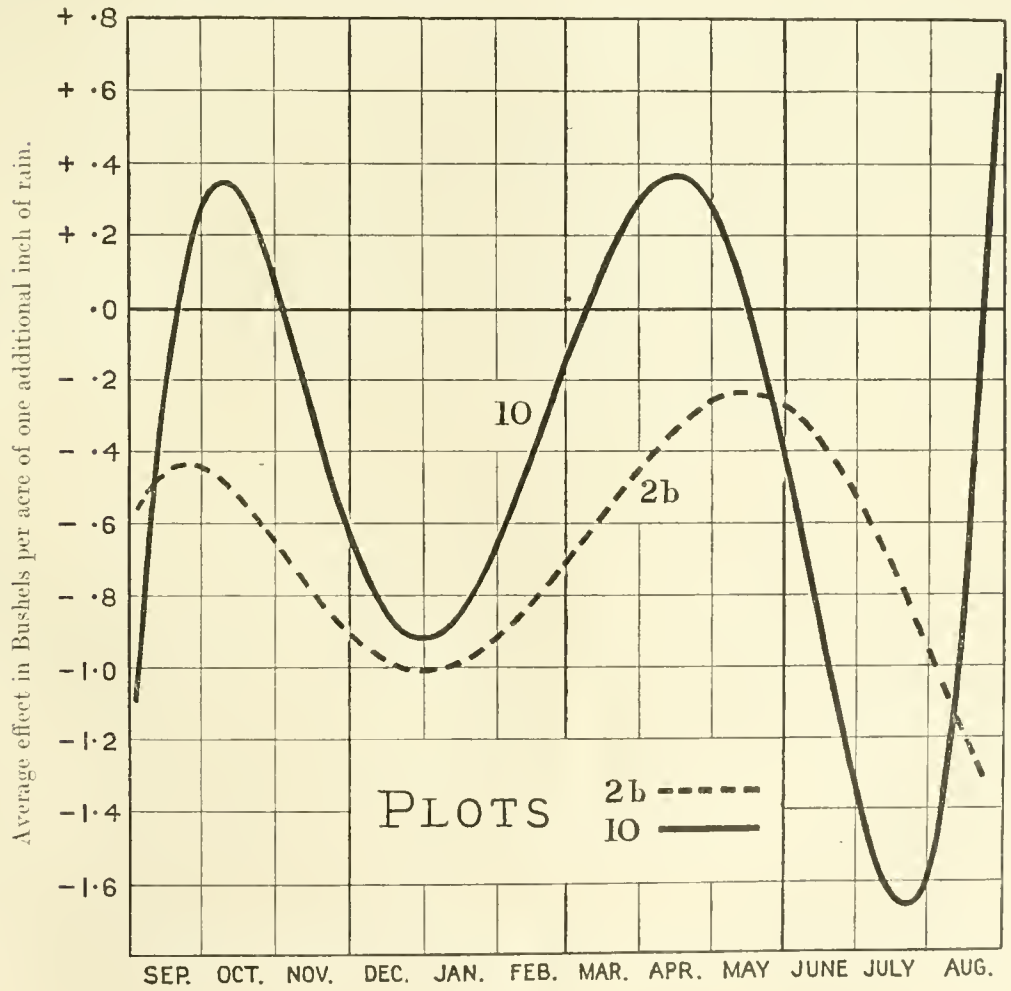



Fig. 8.

In these two plots the yield is very seriously lowered by lack of potash, and in the case of plot 10 by lack of phosphate also. The mean yields are 19·50 and 22·05, with heavy annual deterioration 0·157 and 0·219. Plot 10 with the lower yield is, as might be expected, the more extreme example of the type.

Both curves are positive in October with values + 0·35 (1·8 per cent.) and + 0·16 (0·7 per cent.); the winter damage is less marked in absolute value than in the types III and IV., being only 0·92 (4·7 per cent.) and 0·85 (3·8 per cent.); in April positive values are reached in both cases, those of plot 10 being even higher than in October, + 0·37 (1·9 per cent.) and + 0·03 (0·1 per cent.). The July damage is more extreme than in preceding types being considerably greater than that in January, namely, 1·67 (8·5 per cent.) and 1·43 (6·5 per cent.). In both the April and the July values plot 10 contrasts more strongly than plot 11 with the plots of type IV.

Type VI.—The dunged plot 2B (fig. 8).

This plot differs from all the others in the texture of its soil, and to this perhaps may be ascribed the absence of strong contrasts in the effects of rain at different seasons; the average advantage of dry weather is comparable with the other plots, but the only marked feature of the curve is the severe damage caused by rain immediately preceding harvest. This feature may perhaps be analogous to the heavy damage of July rain in plots 10 and 11, which it should be noted resemble the dunged plot in their relative abundance of nitrogen. Plot 2B differs strikingly from all the other plots, and especially from plots 10 and 11 in the relative constancy of its yield.

The mean yield is 34·55 bushels per acre. The autumn minimum of rain drainage is 0·43 (1·3 per cent.), the winter maximum 1·00 (2·9 per cent.), the summer minimum is well marked and prolonged at 0·24 (0·7 per cent.), after which the curve falls to a final value of 1·39 (4·0 per cent.); as we have seen, the mean damage for harvest rain which catches the crop is no doubt greater.

### 10. *The Value of Rainfall Regressions as Prediction Formula.*

The extent to which the variation in crop yield is predicable from a rainfall record may be calculated from the coefficients of multiple regression of the last section. As we have seen in section 2 the coefficient of multiple correlation gives a much exaggerated notion of the prediction value of a regression formula, if calculated from a small sample. In the present instance, although our series of values is a long one, the number of degrees of freedom has been whittled down to 54 by the rejection of unsuitable years, and the elimination of slow changes; six meteorological variates have been used, which represents the utmost economy in view of the complexity of the meteorological sequences. The distribution of the variance, between the 6 degrees of freedom of the regression formula, and the 48 degrees of freedom in which the variates may differ from the regression formula is shown in Table XIV, in which are also shown the multiple correlation R, and the percentage of variance ($A_1$ of section 2) ascribable to the average effect of rain.

TABLE XIV.

| Plot. | Sum of squares. | | Total 54 degrees. | R. | $A_1$ per cent. |
|---|---|---|---|---|---|
| | Regression formula 6 degrees. | Deviations 48 degrees. | | | |
| 2B | 884 | 996 | 1,880 | 0·6859 | 40·12 |
| 3 & 4 | 149 | 336 | 485 | 0·5548 | 22·13 |
| 5 | 163 | 401 | 564 | 0·5370 | 19·93 |
| 6 | 563 | 1,062 | 1,625 | 0·5885 | 26·16 |
| 7 | 790 | 2,091 | 2,881 | 0·5236 | 18·34 |
| 8 | 959 | 2,083 | 3,042 | 0·5614 | 22·96 |
| 10 | 796 | 1,827 | 2,623 | 0·5509 | 21·63 |
| 11 | 668 | 2,501 | 3,169 | 0·4592 | 11·23 |
| 12 | 1,395 | 2,069 | 3,464 | 0·6315 | 32·79 |
| 13 | 985 | 1,938 | 2,923 | 0·5803 | 25·39 |
| 14 | 938 | 2,069 | 3,007 | 0·5584 | 22·58 |
| 17 & 18 M. | 199 | 686 | 885 | 0·4736 | 12·73 |
| 17 & 18 A. | 965 | 1,189 | 2,154 | 0·6693 | 37·90 |

It will be seen how very inadequate is the value of R to indicate the value of prediction formula ; the extreme values of R in the above totals are 0·459 and 0·686, but these values indicate in this case that in one plot 11 per cent., and in the other 40 per cent. of the variance, is expressible in terms of the sequence of rain records.

It is remarkable that so much of the variance as 40 per cent. should be expressible in terms of a single meteorological element such as rainfall, especially when it is remembered that all causes of variation without exception, including casual errors, and the quadratic terms of the rainfall effect, are included in the remaining 60 per cent. This leads us to think that a record of rainfall, in spite of the many disabilities which have been urged against it, is of more value than the record of any other single element, in characterising the season.

The effects ascribed to rain are in most plots clearly significant ; the values of P, calculated from the formula

$$(1 - R^2)^{24} (1 + 24R^2 + 300R^4)$$

range from 0·0000186 to 0·0659.

The excessive variation of plot 11 thus masks the rainfall effect in the same manner that it masks the slow changes in this plot ; the rainfall effect is somewhat the more important and shows up more clearly. In both cases the similarity of the curves of plot 11 to those of other plots, especially those which it would be expected most nearly to resemble, shows that no serious deviations have been introduced into these by random fluctuations.

The probable errors of random sampling of the regressions of crop on rainfall may be calculated as demonstrated by the author in 1922 (9). The number of years is sufficient to ensure the effective normality of the distributions. In the comparison of the regression, of any two plots, we may anticipate that the random errors are on a substantially smaller scale, since the experience of all the plots is drawn from an identical series of seasons.

No statistical estimate of the accuracy of curves for purposes of comparison can be made in the absence of strictly parallel plots. It may, however, be confidently anticipated that random errors of this kind are, if anything, of a smaller order than the differences between the curves obtained for plots 7, 13 and (17 and 18) ammonium series.

*Reality of Slow Changes.*

The prediction formulæ which we have obtained are entirely independent of the slow changes which appear to have taken place in yield and in weather. In presenting the evidence for slow changes in yield (5) it was presumed, in the absence of a full investigation of the sequence of rainfall records, that favourable and unfavourable weather conditions fluctuated independently from year to year ; in fact, that slow changes were absent from the meteorological series. We now know that a small percentage of the variance of annual rainfall should probably be ascribed to slow changes, and that sequences of wet years did in fact occur about 1879 and 1914. Since these two periods agree with the two main depressions in yield, it becomes at first sight questionable whether the slow changes in yield may not after all be ascribable to meteorological effects.

This question is most readily answered by calculating the actual depression in yield ascribable to the additional rainfall in the two rainy periods. Plot 2B is most suitable for a comparison for the residual variance is, in comparison to the yield, very small in this plot, and the prediction formula correspondingly accurate. The yield in any year is regarded as made of two parts, part (positive or negative) is due to the rainfall sequence under which the crop was grown, the remainder is the yield handicapped for the advantages or disadvantages ascribable to rainfall. The sequence of 10-year means of these two quantities is shown in fig. 9. It will be seen that the yield, after making allowance for rainfall, still shows strongly the slow changes which originally attracted attention. The changes in weather only account directly for a portion of the slow changes observed ; in the first half of the sequence this portion is roughly one-half, but in the latter half it would appear not to exceed one-quarter of the total effect.

The circumstance that in two cases the depression of yield, after allowing for rainfall, coincides with a series of years in which the rainfall was on the average unfavourable, is in accordance with the suggestion previously put forward, that in addition to the immediate effect of rainy weather on the growing crop, a sequence of wet years produced an additional and prolonged unfavourable influence by fostering weeds. Two cases of agreement are, however, quite insufficient to prove that the additional depression of yields is causally connected with the rainy periods. The fact of the coincidence, whether causal or fortuitous, does, however, emphasise the importance of eliminating slow changes in the study of annual figures ; if slow changes had been ignored in carrying through the correlational work, we should no doubt have arrived at greatly exaggerated estimates of the harmful effect of rain. Higher correlations would have been obtained and the results would have exhibited one more case of high but essentially unreliable correlations in meteorological agriculture. Against the view that the depression in yields is an indirect effect of a rainy period must be set the fact that in our diagrams the yield appears

to fall off somewhat *before* the wet sequence has set in ; both series being 10-year means the slight falsification of the sequence incurred by using this form of representation should have a similar effect in both series.

*Effect of Fallow.*

Certain years were rejected owing to the crops grown having followed a fallow ; it was suspected that the yields obtained in these years were unduly high, and would vitiate

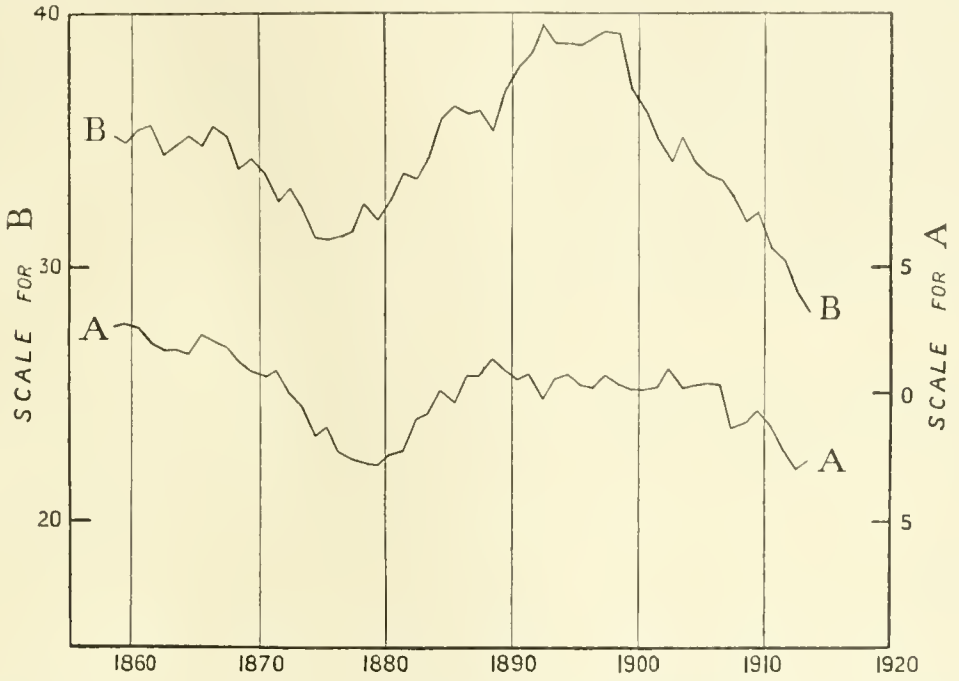



Fig. 9.—Graph A shows the variation in the expectation of wheat yield as judged from the quantity and date of the rain during the year. Graph B shows the additional variation of the actual yields observed after allowance for the expected rainfall effect. Both graphs show 10-year means, in bushels per acre.

the weather correlations if they were retained. A knowledge of the expected effect of weather, together with that of the general fertility of the plot in neighbouring years, makes it possible to estimate the amount by which the fallow benefited the crop. On plot 2B the excess of yield over expectation in 1890, 1891, was 4·48, 6·77 ; in 1905, 1906, the excess values were 6·98, 5·95, while in 1915 the excess was nearly 15·50. In 1916 the two halves of the field were harvested separately, and the half that followed fallow showed an excess of 14·30 bushels over the other. This suggests that the value for 1915 is not so inaccurate as might be thought.

In all these 6 cases the effect of fallow is strongly marked : the average gain is the very large amount of 9 bushels per acre. It should be remembered that this plot receives a heavy dressing of farmyard manure, and it is doubtful if any part of the gain can be ascribed to the accumulation of plant nutrients. The immediate object of fallowing was in all cases the eradication of weeds, and it may be that the increased yield was mainly due to a greater freedom of the land from weed infestation. If that is so the effect must

have been remarkably transient, for the value for 1916 depends on a comparison of the crop immediately following fallow with one which had been fallowed only two years before. The supposition that the benefit was largely a matter of freedom from weeds accords with the higher values of 1915, 1916, when the field, to judge by the average yields on all plots, was much more severely infested than in 1890 or 1905.

The magnitude of the effect of fallow illustrates the complications which arise in the interpretation and statistical reduction of even the best agricultural data. We have attempted to eliminate the large changes which have occurred in the yields of these plots by means of continuous curves, but there is reason to fear that even if the fertility of the soil normally suffers continuous changes from year to year, yet discontinuities were introduced at the periods when fallowing was resorted to. By rejecting the crops immediately concerned the greater part of the errors have no doubt been eliminated, yet the theoretical efficacy of the continuous curve is impaired, however nearly it appears in practice to represent the actual course of events.

### 11. *Comparison with Previous Results.*

In 1880 Lawes and Gilbert published (12) a long and careful account of the meteorological characteristics of seasons favourable and unfavourable to wheat. The object was to effect a qualitative and preliminary enquiry (p. 174): "As yet, however, the connection between meteorological phenomena and the progress of vegetation is not so clearly comprehended as to enable us to estimate with any accuracy the yield of a crop by studying the statistics of the weather during the period of its growth. . . . . But it is only by a careful comparison of the characters of the seasons on the one hand, and of the quantity and quality of the produce on the other, for many years, that we can hope to acquire sufficient knowledge to enable us to assign to the various agencies, the sum of which constitutes the climate of the year, their respective values in the production of the crop."

In respect of rainfall, Lawes and Gilbert, using principally Rothamsted data, but also some instances of years of exceptional harvests prior to their experiments, concluded that comparative dryness was desirable from Seed-time (November) to harvest, especially in Winter and early Spring. Their table of averages shows no exceptional dryness of October for the favourable years, or exceptional wetness in the unfavourable years. The effect of winter and spring rain is ascribed partly to drainage causing loss of nitrates, and partly to hindering root development.

It appears that in the points upon which they laid most stress the conclusions arrived at by the more exact statistical methods now available and with the aid of 39 more seasons' experience at their station, would have caused Lawes and Gilbert only to reaffirm their conclusions more strongly and with greater precision. The comparison of different plots has emphasised the importance of the drainage of nitrates, and points to further influence of manurial conditions on the response of the crop to July rain. The cause of this is at present obscure, but the effect seems well marked. In addition it has

been possible to make some attempt to develop formulæ which shall predict the yield from the weather statistics.

Attention was called to the possibility that autumn rainfall was an important factor in determining the wheat crop by Sir NAPIER SHAW in 1905 (14). SHAW did not use the method of correlation, but pointed out that in the twenty years 1885–1904, with two exceptions, when the yield for Eastern England was above the average the previous autumn rainfall was below the average and *vice versa*. In this set of 20 pairs of values the correlation is actually $-0{\cdot}629$, a value which would only be exceeded by uncorrelated variates in 3 samples of 20 out of a thousand. The particular meteorological variate, autumn rainfall, was picked out of a table giving at least 36 meteorological quantities, and the chance that all of these, if independent and in reality uncorrelated with the crop, should give correlations between $\pm 0{\cdot}629$, was therefore about $0{\cdot}89$. Such a system would therefore be expected to yield so high a correlation only once in 9 trials, and the fact that such a correlation occurred supplied some presumption that autumn rainfall had in reality a perceptible influence on the crop.

In discussing the significance of this result SHAW (14, pp. 318–319) made the interesting suggestion that the association observed might not be due wholly to the effects of autumn weather, but possibly to dry autumns being frequently followed by a favourable succession of weather in Spring and Summer.

> "He would like to say a word with respect of Mr. THOMAS's suggestion that the autumn rainfall was not the dominant factor in determining the subsequent yield of wheat. What surprised him were not the exceptions, but the agreements. Remembering that nine months had to run between the end of autumn and the beginning of harvest, and considering the influence of the intervening rainfall, sunshine and other accidents that might happen to the crop before it was gathered, it was surprising that the connection should be so close as to be expressed possibly numerically. It might be true that two other columns of figures might be tabulated which would show a closer agreement than the two columns put down in the table, but they certainly would not be columns of figures for individual elements. To take two elements out of the whole table, put them side by side, and find them to agree as they did in this case, was astonishing. Of course other influences affected wheat, and it might be that a sequence of influences was required to follow the autumn rainfall in order to bring out the corresponding result. It might be that the relation was a meteorological one, and that a dry autumn itself implied a dry spring or a dry summer, or whatever combination of circumstances was required for a good yield. He could give a certain amount of evidence in favour of the contention that a meteorological relation existed, and that it was not what took place in the autumn alone which might account for the relation, but succeeding events as well, which were associated with a dry autumn. That contingency made the subject one of considerable interest, and one which must be pursued rather more fully than was possible on the present occasion."

A great advance in method is shown by the magnificent paper by Hooker (10) in 1907. Hooker systematically correlated rain and accumulated temperature for 8-week periods throughout the year with the yields of a number of farm crops in Eastern England, using a slightly different area from that used by Shaw. Hooker clearly recognised the two limitations from which the method of correlation suffers, namely that it takes account only of the *linear* relation of the variates, and secondly that the correlations obtained will be much affected by any such meteorological associations between the weather at different times of the year as were suggested by Shaw. It was to remove these two limitations from the treatment of the Rothamsted data that the present method of computing the partial regressions of the yield on the weather at each period of the year was devised, a method which as we have seen can be extended to the discussion of the quadratic terms of the regression function.

Hooker found high negative correlations between the wheat yield and the rainfall for periods centred in October and January, while between them small positive correlations occurred. In May again positive correlations, this time somewhat larger, made their appearance. It is concluded that a dry September-October ranks first among the wheat's requirements, while the effect of winter rain is ascribed partially to the high correlation (+ 0·6) found between the rainfall of weeks 1–8 and that of the preceding weeks 37–44.

In 1922 Hooker (11) gave recalculated figures for the same variates; the results being now based on 35 years, possess considerable significance. Diagram 10 shows the correla-

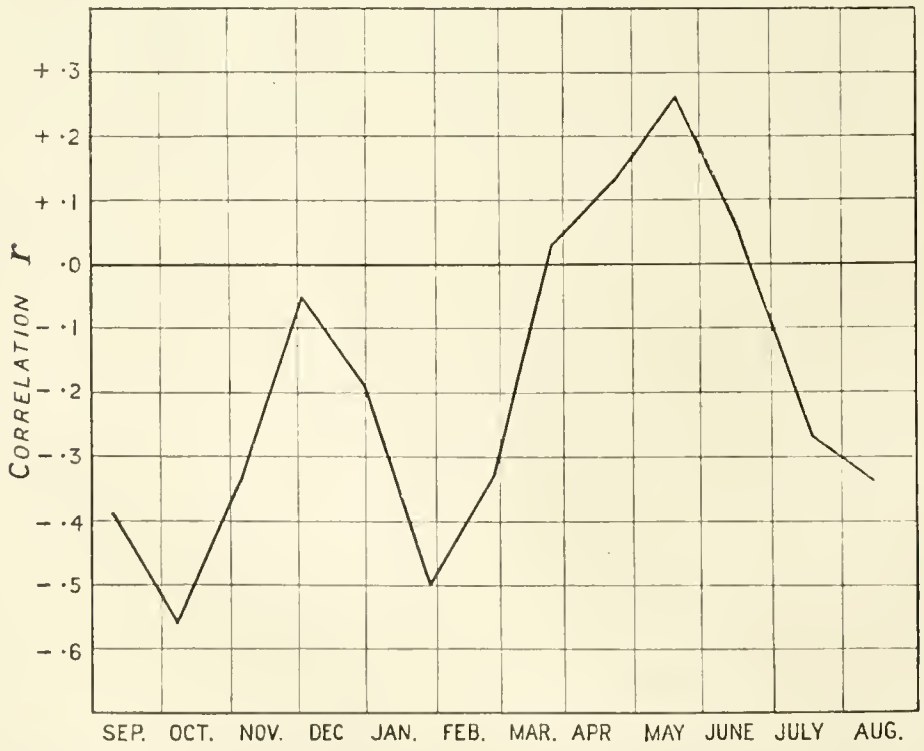



Fig. 10.—Rainfall correlations obtained by Hooker from official returns of wheat yields for Eastern England, using rainfall for 8-week periods centred at the dates shown.

tions obtained for successive (overlapping) 8-week periods, in respect of which HOOKER draws the following conclusions (p. 120) :—

> " Looking first at the curve showing the connection between rain and wheat, the most striking features are the dips about October and in winter, both very nearly identical. As a matter of fact the winter dip is slightly greater than the other. But no weight can be attached to so slight a difference, and we must apparently regard a dry seed-time and a dry subsequent winter as about of equal importance. Sir NAPIER SHAW appears to have been the first to call attention to the great advantage of a dry autumn, and my calculations fifteen years ago pointed to this period being more important than the winter. It is not very clear from the writings of GILBERT and LAWES how far they realised this feature of the wheat's requirements ; it would seem that they were aware of it, but they lay all stress upon the winter rainfall and the washing of the nitrates out of the soil. Wheat is a deep-rooting plant, and I believe it to be probable that, if experiments could be devised to test the point, it would be found that the real effect of a very wet autumn is mechanical—that it is then more the clogging of the soil that prevents the proper development of the root system than the absence of sufficient nitrates. The effect of the loss of the latter is probably more felt later on, during the winter, after the plant is established, and this is reflected in the heavy negative coefficient with rain during the winter months. GILBERT and LAWES have laid so much emphasis upon the washing out of the nitrates that it is generally overlooked that they did in fact recognise the hindrance to root development caused by saturated soil. It seems reasonable to conclude from the data that the mechanical effect of the autumn is just as important as the chemical effect of the winter ; and it is, moreover, one that is practically irremediable."

The statement that the winter dip is slightly greater refers to the partial correlations obtained after eliminating the effects of temperature ; we are here only concerned with the total effects, and it is apparent that HOOKER's data contain substantial evidence for a real deleterious effect of Autumn rainfall. In other respects the series of correlations is not out of harmony with several of the Broadbalk plots, the characteristic period of winter damage is clearly recognisable in both, and we may identify the positive correlations obtained in late Spring with the minimum of damage observed on several of our plots at the same season. The July values also are suggestive of the more highly nitrogenous plots on Broadbalk. The fact that HOOKER's series appears to incline on the whole to more positive values than are observed in our regressions is not improbably due to the inclusion in his data of yields from lighter lands, more susceptible to summer droughts than is the heavy loam of Broadbalk ; in addition his mean rainfall is some four inches less than in our series.

The earlier half of the series, in which the Broadbalk plots are most alike, is alone in contradiction to HOOKER's series. Of course, if strong meteorological correlations existed between the several parts of the year, no resemblance need be expected between

the actual regressions and the correlations which are, in a complicated manner, compounded from them ; and even slight meteorological correlations might seriously disturb the numerical values of the correlations with the crop ; but our meteorological data show that such correlations between weather at different periods of the year, though probably present, are quite small ; and since the agreement, in its general qualitative outline, has not been disturbed from January onwards, it is probable that other factors contribute to the striking difference observable in the effects of October rain. In particular it may be noted that HOOKER's values at this period, allowing for the fact that adjacent points are not independent, but have four identical weeks in common, show somewhat abrupt changes ; the values found for November-December agree well with our maximum in October, and the damage ascribed by HOOKER to "seed-time" may be paralleled by the somewhat increased damage shown in many of our curves in September. In this connection an investigation is desirable of the accuracy with which HOOKER's process of correlating with sets of 8-consecutive weeks, is able to indicate the true maxima or minima of the correlation curve ; and to what errors such estimates are exposed owing to the capricious incidence of heavy rain.

A consideration which more probably contains a solution of the discrepancy is that under industrial farming conditions rain in late autumn and early winter prevents the sowing of large areas of wheat, the land being sown in spring with oats or barley. This is a factor which must strongly influence HOOKER's results, whereas under experimental farming conditions it is inoperative. HOOKER mentions that the wheat area has a correlation $-0{\cdot}41$ with rain of weeks 37–44, but does not mention the correlation of neighbouring periods ; these it would be necessary to know in tracing out the influence of the variable wheat areas on the yield.

It would seem probable that the proportion of land lost to wheat in this way should differ from district to district, and even more probable that such loss should not be proportionately distributed between the two main classes of wheat land in which the wheat follows respectively clover and roots. There are thus many points at which the effect of autumn rainfall upon cropping might give rise to an apparent influence on yield.

It is admittedly by no means necessary that the response of crop to weather should be the same on one particular type of land as it is on the average wheat land of several countries ; the curve of regression may be as much influenced by soil type, as we now know it to be by the manurial condition of the soil. In this respect data of crop averages grown under industrial conditions would be of the greatest value in supplementing by shorter series over a more extensive area the long series from a single field which the Broadbalk data provide. It is to be feared, however, that the Ministry of Agriculture's returns are not sufficiently accurate to meet this need, being based principally upon eye estimates of yield before harvest. Such estimates may be closely correlated with the true yields, and yet fail to give the values of the required regressions even with approximate accuracy ; for there is reason to fear that the deviations from the known mean yield are systematically, though of course unconsciously, underestimated. In

inquiries respecting the weather it would be important to ascertain also that the known meteorological peculiarities of the year under review, the supposed effects of which upon the crops may have been discussed in the press, are without systematic influence upon the judgments of reporters.

## 12. *Summary.*

The study of the theoretical distributions of statistics emphasises the dangers of applying the methods of multiple correlation to small samples, and the necessity of extensive crop data in the study of meteorological agriculture.

By a special procedure involving the analysis of separate meteorological sequences it is possible to obtain an adequate mathematical expression of the average effects of the meteorological influences indicated by different instrumental observations at different times of the year.

The errors involved in the correlation of residuals of series changing in an unknown manner may be minimised by the method of polynomial fitting ; such errors are probably insignificant when this procedure is applied to rain data and wheat yields.

The rain data for Rothamsted have been analysed for 65 years ; there are some indications that the wet years tend to occur in spells ; a continuous and progressive change is observable in the distribution of rain through the year ; in other respects the sequence appears to be fortuitous. Rainfall changes account for only a portion of the slow changes observed in the yields.

Curves showing the average effect on the yield, for each additional inch of rain, throughout the year, have been obtained for 13 plots of Broadbalk wheat field, which have been under uniform experimental treatment since 1852. On all the plots dry weather is generally beneficial. A detailed comparison of the several plots indicates a predominant influence of the effect of rain in removing soil nitrates ; the cause of other well-marked features cannot safely be asserted without further research, which it is hoped may be facilitated by the body of facts expressed in these curves.

Previous investigations bearing particularly on the present data are briefly discussed in the last section.

## 13. *References.*

(1) E. M. Elderton and K. Pearson (1915), "Further Evidence of Natural Selection in Man," 'Biometrika,' vol. 10, pp. 488–506.

(2) E. M. Elderton and K. Pearson (1923), "On the Variate Difference Correlation Method," 'Biometrika,' vol. 14, pp. 281–310.

(3) F. Esscher (1920), "Ueber die Sterblichkeit in Schweden 1886–1914." 'Meddelanden från Lunds Astronomiska Observatorium.'

(4) R. A. Fisher (1915), "On the Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population." 'Biometrika,' vol. 10, pp. 507–521.

(5) R. A. Fisher (1921). "An Examination of the Dressed Grain from Broadbalk." 'J. Agricultural Science,' vol. 11, pp. 107–135.

(6) R. A. Fisher (1921). "On the 'Probable Error' of a Coefficient of Correlation deduced from a Small Sample." 'Metron,' vol. 1, Part 4, pp. 1–32.

(7) R. A. Fisher (1922). "On the Mathematical Foundations of Theoretical Statistics." 'Phil. Trans.,' A, vol. 222, pp. 309–368.

(8) R. A. Fisher and W. A. Mackenzie (1922). "The Correlation of Weekly Rainfall." 'Q. J. R. M. S.,' vol. 48, pp. 234–245.

(9) R. A. Fisher (1922). "The Goodness of Fit of Regression Formulæ, and the Distribution of Regression Coefficients." 'J. R. Statistical Society,' vol. 85, pp. 597–612.

(10) R. H. Hooker (1907). "Correlation of the Weather and Crops." 'J. R. Statistical Society,' vol. 70, pp. 1–42.

(11) R. H. Hooker (1922). "The Weather and the Crops in Eastern England, 1885–1921." 'Q. J. R. M. S.,' vol. 48, pp. 115–138.

(12) J. B. Lawes and J. H. Gilbert (1880). "Our Climate and our Wheat Crops." 'J. R. Agricultural Society,' Second Series, vol. 16, pp. 173–210.

(13) W. T. Russell (1922). "The Relationship between Rainfall and Temperature as shown by the Correlation Coefficient." 'Q. J. R. M. S.,' vol. 48, pp. 225–230.

(14) W. N. Shaw (1905). "Seasons in the British Isles from 1878." 'J. R. Statistical Society,' vol. 68, pp. 247–313.

(15) W. F. Sheppard (1912). "Reduction of Errors by Means of Negligible Differences." 'Fifth International Congress of Maths., Camb.,' p. 348.

(16) "Student" (1917). "Tables for Estimating the Probability that the Mean of a Unique Sample of Observations lies between $-\infty$ and any given Distance of the Mean of the Population from which the Sample is Drawn." 'Biometrika,' vol. 11, pp. 414–417.

(17) A. Walter (1910). "The Sugar Industry of Mauritius." A. L. Humphries, London.

# IX. *The Interaction of Clay with Water and Organic Liquids as Measured by Specific Volume Changes and its Relation to the Phenomena of Crumb Formation in Soils.*

*By* E. W. Russell, *M.A., Ph.D.*

(*Soil Physics Department, Rothamsted Experimental Station, Harpenden.*)

(*Communicated by Sir* John Russell, *F.R.S.*)

(Received December 29, 1933.—Read April 26, 1934.)

The densities of clay particles determined by the specific gravity bottle method vary somewhat according to the liquid used. Thus water, alcohol, and benzene all give slightly different values. The results show that some condensation of the liquid on the surface of the particles usually, if not invariably, takes place, indicating a certain amount of interaction between the clay and the liquid.

The investigation described in this paper shows that this interaction depends partly on the exchangeable bases associated with the clay, and partly on the molecular constitution of the liquid. The relations are as follows :

(1) No evidence could be obtained of interaction between clay and non-polar liquids.

(2) Interaction took place in all the polar liquids examined. Its amount was measured by the reduction in specific volume of the clay as compared with its volume in tetralin, a very convenient non-polar liquid for specific gravity determinations. This reduction in specific volume is approximately proportional to the number of exchangeable cations present, figs. 10, 11, and to the mean density of their surface charge, Table IV. Thus the order of effectiveness of the cations is : magnesium—most active, calcium, hydrogen, sodium, potassium—least active. This order is independent of the liquids used provided no chemical action takes place between the exchangeable cation and the liquid.

(3) A similar effect is produced by some other property of the clay associated with the shape of the titration curve.

(4) The amount of interaction depends on the liquid, but no simple relations with the molecular constitution have yet been discovered.

(5) During the process of removing the liquid by drying, the clay particles may aggregate to form definite crumbs.

 [Published August 14, 1934.

*The Phenomena of Crumb Formation*

Considerable attention has been paid to crumb formation because of its importance in natural soils. Good tilth, which is essential for successful plant growth, is largely dependent on a stable crumb structure of the soil, and much of the art of cultivation consists in the production of this structure prior to the sowing of the seed. Considerable success has been attained by purely empirical methods, but hitherto no satisfactory explanation of the mechanism of crumb formation has been offered.

Crumb particles, similar to those found in nature, can be produced artificially in the laboratory by mixing clay with water and allowing it to dry under definite conditions. These crumbs are characterized by two distinct properties: they are hard when dry and they possess fairly characteristic shapes. The following are the main phenomena of crumb formation and stability brought out in the present investigations.

(1) Crumb formation is limited to those particles possessing considerable base exchange capacity. It is not shown by any of the kaolinitic clays of low base exchange capacity so far examined.

(2) Crumb formation is further limited to particles below a certain size. All the fine clay particles formed aggregates, though no fine clay has as yet been found with a low base exchange capacity. Silt particles of settling velocity exceeding 3·5 $\mu$/sec. did not form aggregates however large their base exchange capacity might be.

The coarser clay minerals and the fractions devoid of base exchange are thus not essential for aggregate formation, but seem to serve only as a diluent, keeping the aggregates more open and mechanically weaker.

(3) Crumb formation is induced only by those liquids the molecules of which have an appreciable dipole moment. Thus water, methyl, ethyl, and amyl alcohols, and nitrobenzene, all gave definite aggregates, while benzene and carbonate tetrachloride gave no real aggregates at all.

(4) The hardness of the crumbs is dependent on the sizes of the clay particles, of the exchangeable cations and of the molecules of the wetting liquid. If any of these be large the aggregates formed are soft: only when they are all small are the aggregates hard. Thus, soft aggregates are formed from coarse clay particles, however large their base exchange capacity; from clays holding large complex cations such as phenyl-benzyl-dimethyl-ammonium; and from clays of small particle size and small cations wetted with liquids like nitrobenzene or amyl alcohol, although these same clays if wetted with methyl or ethyl alcohol form hard aggregates.

(5) The process of crumb formation may be reversible or irreversible; in the former process the crumbs fall down to a paste on rewetting, in the latter they do not, but remain stable. The stability depends on the liquid and on the exchangeable cations of the particles. Sodium clay aggregates are unstable in water, though stable in dry amyl alcohol. Calcium clay aggregates, on the other hand, are stable in water though unstable in dry amyl alcohol.

The formation of crumbs, therefore, appears to depend on an interaction between the exchangeable cations of the clay and the dispersion liquid.

*The Mechanism of Crumb Formation*

Clay particles suspended in water possess an electrokinetic potential and contribute, as SCHOFIELD* has shown, to the electrical conductivity of the suspension. This indicates that each particle is surrounded by an electrical double layer, the outer one being diffuse and consisting of cations, while the inner layer consists of negative charges presumably anchored on the surface of the particle. The cations in the diffuse layer move about in the water in the same way as they do around a complex anion, as pictured in the Debye-Hückel theory of strong electrolytes. Since the water molecules possess a dipole moment, they tend to be orientated along the lines of electric force radiating from each ion in the diffuse layer and from each free charge on the surface of the clay particle. Every cation and particle is thus surrounded by an envelope of orientated water molecules, and the orientation manifests itself as an apparent adsorption or immobilization of water by the clay. Some of the water molecules near an electric charge may be so strongly orientated that they appear bound to it, their heat motion merely making them oscillate about the lines of force, while those further away from the charge possess only a statistical orientation around it. A clay particle in a dilute suspension can, therefore, be pictured as consisting of a central core surrounded by a surface carrying a negative charge. Around each negative charge is an envelope of water molecules which are more strongly orientated the nearer they lie to the charge. Outside this surface are the cations, also possessing envelopes of orientated water molecules. Some cations are so close to a negative charge on the surface of the clay particle that the two water envelopes belonging to these charges overlap and the water molecules in this region are orientated in their joint field. This orientation is very strong, since the negative end of the water dipole is attracted to the cation and the positive end to the particle's surface, as shown diagrammatically in fig. 1.

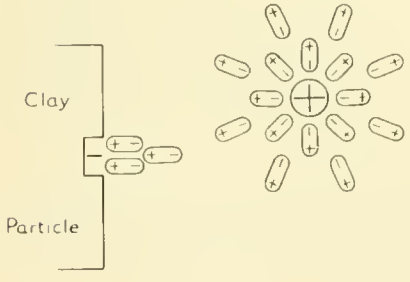



FIG. 1.

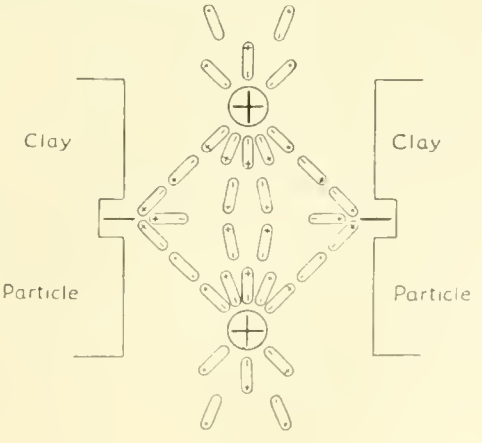



FIG. 2.

* Unpublished data.

As the water is removed the deflocculated clay suspension becomes more concentrated and an increasing proportion of the water molecules become orientated in the joint field of a positive and a negative charge. It is reasonable to assume that a certain number of cations will share their orientated envelopes with two clay particles, as is shown diagrammatically in fig. 2.

A linking system is thus set up consisting of :—

particle—orientated wetting molecule—cation—orientated wetting molecule—particle.*

As removal of water proceeds, an increasing proportion of cations share their water envelopes with two clay particles, and so the number of links increases. The links also become stronger, because they become shorter and the mean orientating couple on the water molecules increases. In consequence, the cohesion of the clay particles, *i.e.*, the hardness of the crumbs, increases.

The hypothesis involves two assumptions, which have not yet been directly tested, though there is some indirect evidence in their favour :—

(1) that a system of the type postulated is in fact mechanically stable ;

(2) that the number of cations dissociating from a clay particle in a system that is losing water decreases less rapidly than the number of water molecules.

The hypothesis accounts for the formation of crumbs by other polar liquids as well as water, and for their non-formation by non-polar liquids. It also explains the necessity for the clays to possess dissociating cations, *i.e.*, base exchange capacity.

Further, it explains the observed relation between the hardness of the aggregates and the size of the clay particles, the cations and the molecules of the wetting liquid. For the smaller the clay particles, the larger the number of links per unit volume between them, provided the surface of the clay particle be curved and not plane. Again, a large univalent cation, such as the phenyl-benzyl-dimethyl-ammonium ion, has a much smaller surface charge density than a potassium ion, and so forms weaker dipole-cation-dipole links and therefore weaker aggregates than those of potassium clay. Finally, the strength of the bond between cation and clay particle increases with the dipole moment of the liquid, provided the position of the dipole in the molecule remains about constant. But if one end of the dipole becomes large, as for example in amyl alcohol and nitrobenzene where the positive ends are the groups $C_5H_{11}$ or $C_6H_5$ while the negative end is the simple group -OH or $-NO_2$, then the positive end of these dipoles cannot be strongly attracted to the negative charges on the clay surface, and so cannot form a strong binding link. This is in agreement with the experimental result that hard crumbs can be formed from water and ethyl alcohol, when both ends of the dipole are small, but not from amyl alcohol or nitrobenzene.

Crumb formation is thus seen to occur when each end of the dipoles of the molecules of the wetting liquid is reactive, so that one end interacts with the negative charge on the clay particle and the other with the positive charge on the dissociated cation. If,

* For convenience this is called in later paragraphs the dipole-cation-dipole link.

however, only one end of the dipole is strongly reactive, then hard crumbs are not formed though considerable interaction with the clay may still take place as shown by the reduction in specific volume. For example, clays interact as strongly with amyl alcohol as with ethyl alcohol, and with considerable diminution in specific volume in each alcohol, yet they form hard crumbs from ethyl and not from amyl alcohol.

*The Influence of Electrolytes on Particle Cementation: Relation of Crumb Formation with Floeculation*

The phenomena of crumb formation on drying after wetting with water are modified in the presence of electrolytes.

Two effects are produced :—

(1) the crumbs are smaller and weaker ;

(2) provided the concentration of the electrolyte exceeds a certain value all crumbs are stable in the solution of the electrolyte.

Both may be attributed to the reduction in dissociation of the cations from the clay surface caused by the electrolyte.

This reduces the number of dipole-cation-dipole links producible on drying and therefore weakens the resulting crumbs.

While crumb instability has not yet been studied in detail there is reason to connect it with the dissociation from the clay particles of a large number of cations having voluminous hydration or solvation envelopes. The presence of the electrolyte, by decreasing both dissociation and hydration of the cations, therefore increases the stability of the crumbs.

Analogies have often been traced between floeculation (*i.e.*, floc formation) in a clay suspension and crumb formation in a drying soil ; actually these analogies are not valid, the processes being different. Floc formation has recently been discussed by the author* for aqueous solution and by McDowell and Usher† for organic liquids. It occurs only when the electrokinetic potential of the clay particle is low, *i.e.*, when only small dissociation of the cations occurs. Crumb formation, on the other hand, occurs when there is an appreciable amount of dissociation of cations from the surface, *i.e.*, when the electrokinetic potential is high.

Flocs have a very open heterogeneous structure and a large volume in comparison with that of their component particles : they are weakly aggregated and can be broken up by shaking. Crumbs, on the other hand, have a much closer, more homogeneous structure, and can be much more strongly held together.

No satisfactory hypothesis has been put forward to account for the cohesion of clay particles into flocs and therefore no detailed comparison with the mechanism of crumb formation can be made.

* ' J. Agric. Sci.,' vol. 22, p. 165 (1932).

† ' Proc. Roy. Soc.,' A, vol. 131, pp. 409, 564 (1931).

The essential difference between the two processes is brought out by comparing the effect of adding calcium chloride to a 5% suspension of calcium-clay in water, with the effect of adding sodium chloride to a 5% suspension of sodium-clay. The calcium-clay forms flocs which settle to a fairly loose sediment, but the sodium-clay forms no flocs while the whole suspension sets and becomes rigid and thixotropic. The calcium chloride strongly depresses the dissociation of calcium ions from the clay surface, so that floc formation can take place, while sodium chloride cannot depress the dissociation of sodium ions from the clay to anything like the same extent, so that floc formation cannot take place. The sodium and chloride ions introduced by the added salt, as well as the sodium ions that have dissociated from the clay, have envelopes of orientated water molecules around them, and so can form links of the type

$$\overset{+}{(\text{dipole})}\ \overset{-}{(\text{cation})}\ \overset{+}{(\text{dipole})}\ \overset{-}{(\text{anion})}\ \overset{+}{(\text{dipole})}\ \overset{-}{(\text{cation})}\ \overset{+}{(\text{dipole})}$$

which can bind two clay particles together. This link is of the same type as the simple dipole-cation-dipole link illustrated in fig. 2, but it is longer, it can hold much more water and therefore becomes effective at much higher water concentrations. The length is responsible for the thixotropic behaviour of these rigid suspensions, since if the system is violently disturbed the binding links are broken down and take a certain time to build up again.

## *Experimental*

*Materials*—Clays separated from the following soils were used in this investigation.

*Alluvial or drift soils*—Clay-with-flints (Rothamsted), Boulder clay (Samuelstown, Haddington), three London Clays (Whitstable, Herne Common and Chessington, near Surbiton), Carse soil (Larbert), Fen soil (Cambridgeshire), Gezira soil (Wad Medani, Sudan) and a Palestine soil (Mikveh-Israel).

*Soils from solid formations*—Old Red Sandstone (Putley, near Ledbury), Triassic Keuper Marl (Long Ashton), Lower Lias (Axminster), two lower Oolitic soils (Andoverford, Gloucester, and Redbourne Kirton-in-Lindsey) and an Oxford clay (Bampton Lew).

To prepare the separated clays, the soils were usually first treated with sodium hypobromite to oxidize and to bring into solution much of the organic matter, and to assist in the breaking up of the soil crumbs. The soil was then repeatedly dispersed in a large volume of dilute (0·05%) sodium carbonate. The material that remained in suspension in the top 20 cm after 24 hours' standing was then syphoned off and flocculated with dilute acid. The bases were then washed out, with dilute HCl (less than N/50) in the first series of separations, and with half-neutralized sodium acetate ($p_{\text{H}}$ about 4·5) in the later series, until the supernatant liquid was calcium-free; the acid or sodium acetate was then washed out with distilled water until the clay was deflocculated. In the first series of experiments the clay was divided into aliquots, and each was saturated with a given base by adding the corresponding acetate. The

acetates were added repeatedly until the $p_H$ of the supernatant liquid was about 7, when the alkaline-earth acetates were washed out with distilled water, and the alkali acetates with distilled water followed by alcohol. In the second series of experiments the purified hydrogen or sodium clay was passed through a Sharples supercentrifuge to remove the coarse clay and fine silt particles present, so that the deflocculated clay suspension would not settle appreciably during preparation, thus ensuring that the crumbs produced should have as uniform composition as possible. Instead of preparing the different homionic clays—that is, clays containing only one exchangeable base—by adding the acetate, either of two buffer solutions were used : half-neutralized *p*.nitrophenol—$p_H$ about 7, and half-neutralized acetate—$p_H$ about 4·5, in each of which the base concentration was N/50.*

The liquids used for wetting were, whenever possible, of A.R. standard of purity, and were always dried either over phosphorus pentoxide or over freshly ignited lime for 3–4 weeks and distilled just before use. No more refined drying technique was employed, as the liquids were in contact with the air for short periods and so could absorb traces of water vapour just before and just after wetting the crumbs.

### *The Experiments on Crumb Formation and Crumb Stability*

(*a*) *Relation of particle size to hardness of aggregates*—Silt fractions were separated from the Carse, Lower Lias and Gloucester soils, and coarse clay fractions from the Gloucester and one of the London clay soils. The coarser materials were easily separated from the silt and clay by sedimentation, the silt used having a settling velocity between 20 and 3·5 μ/sec. The fine clay crumbs were dispersed by repeated alternate treatment with sodium hypobromite and boiling with hydrogen peroxide, and the clay was removed by repeated dispersion in sodium carbonate. These treatments were continued until no appreciable amount of clay came into suspension. The coarse clay fractions were prepared from the complete clay fraction by passing the deflocculated clay through a Sharples supercentrifuge, removing the coarse fraction from the bottom of the bowl and repeatedly passing this fraction through the centrifuge to remove the finer clay particles from it. The settling velocities of these clay particles cannot be calculated from their method of preparation, nor has an ultra-microscopic count been made on the dispersion, so that it is only possible to give the upper limit of their settling velocities, namely, 3·5 μ/sec. The base holding power of these materials between Schofield's† hydrochloric acid level and his *p*-nitrophenol level are in milli-equivalents/100 grams :—

For the silts : Carse 2·2, Lower Lias 3·5, Gloucester 10·8 ;

For the coarse clays : Gloucester 33·5, London I 23·2.

* The author is indebted to Dr. R. K. Schofield for bringing to his notice the advantage of using strongly buffered solutions for the preparation of homionic clays, and for suggesting the most appropriate weak acids to use.

† 'J. Agric. Sci.,' vol. 23, p. 252 (1933).

On drying, or on freezing and drying from a water dispersion, these materials formed weak blocks. The silts broke down immediately to a fine powder when touched with the finger, while the coarse clays had to be rubbed between the fingers.

(*b*) *Relation of size of the exchangeable ion to hardness of crumb.*—Two separate fractions of a Gloucester clay, a medium fine and a very fine fraction, were divided into five aliquots, which were saturated up to $p_H$ 7 with Ca, Na, K, Mg and phenyl-benzyl-dimethyl-ammonium. The phenyl-benzyl-dimethyl-ammonium chloride was prepared by adding equimolecular quantities of dimethyl aniline and benzyl chloride in absolute alcohol, leaving in the dark for one month and precipitating the salt with ether. The crystals were then washed with ether, redissolved in alcohol, reprecipitated and rewashed with ether and dried. The acetate or chloride was then washed out, and the deflocculated clay paste either frozen or, if it was the sodium or potassium clay, dried. For each of the calcium and magnesium clays hard crumbs, diameter about 3 mm, were formed when dry, but for the substituted ammonium clay the crumbs were smaller and crumbled between the fingers to a fine powder ; this clay held nearly 60 milli-equivalents of exchangeable base per 100 grams, and fully deflocculates only with difficulty. When dry it has a very faint smell of dimethyl aniline.

(*c*) *Relation of wetting liquid to hardness of crumbs*—Clay dispersions made up in different liquids were obtained from the deflocculated clay in water by removing the water with alcohol, the alcohol with the required liquid or with benzene, or the benzene with the required liquid. At each renewal of the liquid the clay was shaken for several hours to ensure good dispersion. The liquid was removed with Berkefeld filter candles attached to a water pump through sulphuric acid bottles, only dry air being admitted to the bottles containing the clay. Freshly distilled dry liquids were always used finally. These clay dispersions were then dried at 80° C over $P_2O_5$ in a desiccator kept at a low pressure with an oil pump containing an ice condenser and silica-gel adsorption tubes in the vacuum circuit. Hydrogen-clays separated from the Old Red Sandstone, Lower Lias and London Clay soils, and a calcium London clay gave only extremely soft aggregates from benzene and carbon tetrachloride, not quite so soft from amyl alcohol, and hard from ethyl alcohol dispersions, while definite, though fairly soft, crumbs were formed by freezing and drying a nitrobenzene dispersion of the hydrogen Lower Lias clay.

(*d*) *Relation of wetting liquid to stability of crumbs*—If a crumb formed by freezing a water paste of a calcium clay is thoroughly dried and then wetted with dry amyl alcohol under vacuum it usually, though not always, broke down into a paste, but no sodium clay crumb broke down. It was also found that a deflocculated suspension of a calcium clay can readily be formed in dry amyl alcohol, while it has not yet been possible to deflocculate a sodium clay in it.

### *Determination of the Specific Volume of the Clays*

To determine the specific volume of the clays, about 10 grams of clay crumbs were put in a 25-cc specific gravity bottle, placed over $P_2O_5$ in a desiccator, which was

evacuated by the oil pump, and left for a week. Fresh $P_2O_5$ was then put in, the desiccator re-evacuated and left for at least two weeks, and this stage would be repeated unless the surface of the $P_2O_5$ was still powdery. For non-volatile liquids wetting was carried out under vacuum with the oil pump running, but for volatile liquids, such as water or the alcohols, this method was not satisfactory, since the clay crumbs would be shot out of the bottle as the liquid came into contact with them. For volatile liquids two different methods of wetting were used. In the first, after the crumbs had been evacuated for about 20 minutes, the vacuum circuit was shut and the liquid was run into the bottom of the desiccator, and after this had been left for about 10 minutes fresh liquid was then run into the bottles containing the crumbs. This method was still not very satisfactory as some bumping often occurred when the liquid first came in contact with the clay. In the second method tried, the desiccator was evacuated, then a stream of the vapour was passed over the crumbs to enable them to adsorb an appreciable amount of the liquid, then the liquid itself was run into the bottles. This method, which was used only at the end of the experiments, was much more satisfactory. After the clay had been wetted under vacuum, the bottles were put into closed tanks containing dry paraffin for 3–6 hours, these tanks being immersed in a large water thermostat tank running at about 25° C. Three hours was found to be sufficiently long to allow the clay to reach equilibrium with the liquid.

In the calculation of the apparent specific volume of the clay crumbs the density of water was always assumed to be that given in Landolt-Börnstein's tables, all other densities being calculated from this. Buoyancy corrections were applied to the weights of the clay and of the liquids, the air density being determined each time from the barometer, thermometer and wet and dry bulb readings with the use of Landolt-Börnstein's tables. The differential buoyancy correction needed for the layer of adsorbed liquid was not large enough to be significant. The error introduced in the determination of the specific volume of the clays under the experimental conditions, due to limitations in the accuracy of the instruments used, is about $10^{-4}$, this being caused by errors of 0·1 mg in the weighings and 0·05° C in reading the temperature in the thermostat, these two sources of error being about equal. The only large sources of error are loss of clay from the bottles during wetting under vacuum, entrapping of air, and loss of liquid by evaporation between taking the bottle out of the thermostat and weighing it. Each of these three sources of error will make the specific volume of the clay appear too large, and as a matter of fact no case has yet arisen where the specific volume appeared to be obviously too low. Duplicate determinations were not made together, and everything possible was done to make the maximum amount of systematic error occur as random error between duplicates. From a study of the actual deviations between 400 duplicate determinations of the specific volume of the clays available, only 43% had divergences less than $2·5 \times 10^{-4}$, in 25% it was between $2·5–5·0 \times 10^{-4}$, in 12% between $5–10 \times 10^{-4}$, and in 20% it exceeded $10^{-3}$. Thus in 20% of the duplicate determinations at least one of the determinations had an error of $10^{-3}$, so that

at least 10% of the individual determinations had to be rejected owing to one of the three major sources of error, of which loss of clay during wetting under vacuum was the most common. It thus appears probable that the error of the mean is usually about $2 \cdot 5 \times 10^{-4}$. It is possible to make an independent estimate of this error, for some of the determinations were repeated at different times, and by comparing the differences between their means, approximately the same figure is obtained. In conclusion, therefore, the error in a difference between means will be about $5 \times 10^{-4}$.

### *Specific Volume of Clays in Different Liquids*

Few of the methods devised in general colloid science to measure the interaction of a liquid or gas with a solid substrate have been, or can be, applied to clays, and so far no systematic investigation of this interaction in clays has been undertaken. The method adopted by the author has been to measure the apparent density of carefully dried soil and clay crumbs in different liquids. This method, which was first used by Williams* and Cude and Hulett† for charcoal, seems to have been used for soils only by Anderson and Mattson,‡ who found that dried separated clays appeared to be heavier in water than in toluene, and by Tschapek§ who found that soils appeared heavier in water than in the four hydrocarbons benzene, xylene, petrol, and petroleum (? paraffin), in which the density was about the same. The method has the disadvantage that the results are not easy to interpret, since four separate factors can influence them :—

(1) the interaction of the particle surface with the dispersion medium ;
(2) the proportion of the total surface area of the clay particles forming the crumbs accessible to the wetting liquid ;
(3) the compression of the liquid in the fine capillaries of the crumbs due to the large curvature of the solid-liquid interfaces ;
(4) the amount of adsorbed and entrapped gases in the crumbs after wetting.

The adsorbed and entrapped gases are usually removed by a strong outgassing by keeping the system at a high temperature and a very high vacuum for a considerable time, but this method is probably inapplicable to clays as certain irreversible processes seem to attend high temperature dehydration, 100-200° C, though this point has never been adequately examined. For this reason, and also because the work was largely exploratory, no elaborate degassing technique was employed, wetting being carried out slowly in a desiccator connected to a small rotary Hyvac oil pump.

The experiments were designed to show the influence of the exchangeable ions, held by a clay, on the apparent specific volume of the clay when immersed in different

* ' Proc. Roy. Soc.,' A., vol. 98, p. 223 (1920).
† ' J. Amer. Chem. Soc.,' vol. 42, 391 (1920).
‡ ' U.S. Dept. of Agric. Bull.,' 1452 (1926).
§ ' Kolloidzschr.,' vol. 63, p. 34 (1933).

liquids. Separated clays were used, as very uniform crumbs can be obtained from them, so that sampling errors are negligible compared with other sources of error. This condition can barely be obtained with soils, for during the preparation of a soil holding one metallic cation only, the coarse and fine fractions become partially separated, through the crumbs breaking down in the solutions used, and it is very difficult thoroughly to remix all the materials. Since it was also considered more important to use only a few liquids but as many clays as possible, only two liquids, water and tetralin (tetrahydronaphthalene), were used for every clay, though amyl alcohol, aniline and nitrobenzene were also used for most of them. Isolated experiments in a number of other liquids were also made. Considering these five liquids only, the specific volume of a clay was greatest in tetralin and next in nitrobenzene without exception. With the exception of some of the hydrogen clays, the specific volume in aniline was greater than in water, and greater in amyl alcohol than in water, with the exception of the magnesium and one of the calcium clays, and about the same in amyl alcohol and aniline for the sodium, potassium and most of the calcium clays. Since the specific volume of a clay is greatest in tetralin, these data will first be examined to see how far they support the hypothesis that tetralin has no interaction with the exchangeable ions on the clay.

*The specific volumes of clays in tetralin*—The specific volume of a clay in a liquid depends on the following factors :—

(*a*) the specific volume of the clay substrate ;
(*b*) the weights and volumes of the exchangeable ions on the clay ;
(*c*) the interaction between the exchangeable ions on the clay particles and the wetting liquid ;
(*d*) the interaction between the clay substrate and the wetting liquid ;
(*e*) for dried clay crumbs the sizes of the pores in the clay crumb, these determining both the accessibility of the liquid molecules to the clay surface, and the compression of the liquid in the pores due to interfacial tension forces.

By considering the apparent specific volumes of different ionic clays in a given liquid, it should be possible to decide if the variations between the specific volumes of a clay holding different exchangeable ions can be accounted for completely by the weights and volumes of the exchangeable ions ; if so, the exchangeable ions would have no differential effect on the total volume of the fine pores in the crumbs, nor any differential interaction with the wetting liquid.

The contribution of the exchangeable ions to the specific volume of the clay can be calculated as follows. Let $\sigma_x$ and $\sigma_y$ be the specific volumes of a clay holding B equivalents of ion X and ion Y respectively in a given wetting liquid, and let V, $V_x$ and $V_y$ be the volumes of one equivalent of clay, ion X and ion Y respectively and $M_x$ and $M_y$ the equivalent weights of the two ions, then

$$\sigma_x = \frac{V + BV_x}{1 + BM_x}, \text{ and } \sigma_y = \frac{V + BV_y}{1 + BM_y}$$

or, on eliminating V

$$\sigma_y = \sigma_x \left(\frac{1 + BM_x}{1 + BM_y}\right) - B\left(\frac{V_x - V_y}{1 + BM_y}\right).$$

The two denominators of this equation can be expanded, and all except the first two terms of each expansion ignored since $BM_y$ does not exceed 0·03 in any of the clays considered, giving as an approximate equation

$$\sigma_y - \sigma_x = B\sigma_x(M_x - M_y) - B(V_x - V_y). \quad \ldots\ldots\ldots (1)$$

Unfortunately, although $M_x$ and $M_y$, the equivalent weights, are known, $V_x$ and $V_y$, their equivalent volumes, are not, so that this equation cannot be used to calculate $\sigma_y - \sigma_x$. Using the experimentally determined values of $\sigma_x$ and $\sigma_y$ for a series of clays, it is, however, possible to examine whether $V_x - V_y$, calculated from this equation, is independent of the type of clay used, and so dependent solely on the ions X and Y, and whether the observed values of $V_x - V_y$ appear reasonable on other grounds. The results given in table I, show that $V_x - V_y$ is nearly independent of the clay type, though, owing to the large experimental errors attached to this difference, it is not possible to say if the values for individual clays really differ significantly. Table I also shows that, per equivalent, the ionic volume of the ions increase in the order sodium, hydrogen, calcium and potassium, magnesium.

It is possible to obtain further evidence for the validity of these results from an analysis of the differences in molecular volume between different crystals belonging to the same system. For example, the densities of sodium chloride and potassium chloride crystals are given in the International Critical Tables as 2·163 and 1·988, and both belong to the cubic system. Hence, one mole of sodium chloride occupies 27·0 ccm and of potassium chloride 37·5 ccm, so that one mole equivalent of potassium is 10·5 ccm larger than one equivalent of sodium, on the assumption that the whole difference in volume between the sodium and potassium crystals is taken up by the larger size of the potassium ion, and ignoring all volume changes due to differences of packing. This calculation can be made for a large number of similar sodium and potassium crystals, and for the six available monovalent anions tabulated in the International Critical Tables the value of $V_K - V_{Na}$ varies from 12·6 to 10·3, with a mean of 11·3 ccm, though for the divalent anion crystals the figure is rather lower. This is very close to the value 11·1 ccm determined from the specific volume of the clays. The difference between the calcium and magnesium ions is more difficult to determine since the crystals of the corresponding salts usually belong to different systems, though it seems to vary from about 6—11 ccm with a mean of about 8 ccm per mole, so that $V_{Ca} - V_{Mg} = 4$, and from the scanty data available $V_{Na} - V_{Ca}$ is about 3 ccm. Since the results from the density calculation are $V_{Mg} - V_{Ca} = 14·6$ and $V_{Ca} - V_{Na} = 10·0$, these two sets of results are neither in quantitative nor qualitative agreement. Thus, while the difference between the volumes of sodium and potassium ions on a

TABLE I—Calculated Difference in Specific Volumes of Different Ions on Clays

| | Bases held milli-equiv/ 100 grams | Ca — Na. | | | H — Na. | | K — Na. | | Mg — Ca. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Delta\sigma$ (obs.) $\times 10^{-5}$ | $V_{Ca} - V_{Na}$ cc/equiv | $\lambda_{Ca} - \lambda_{Na}$‡ | $\Delta\sigma$ (obs.) $\times 10^{-5}$ | $V_H - V_{Na}$ | $\Delta\sigma$ (obs.) $\times 10^{-5}$ | $V_K - V_{Na}$ cc/equiv | $\Delta\sigma$ (obs.) $\times 10^{-5}$ | $V_{Mg} - V_{Ca}$ cc/equiv | $\lambda_{Mg} - \lambda_{Ca}$‡ |
| Trias ... ... ... | 20·2 | 125 (10)* | 5·1 | 0·92 | 250 (13) | 4·5 | 77 (32) | 9·5 | 312 (6) | 12·7 | 1·80 |
| Rothamsted ... ... | 22·7 | 303 (45) | 12·2 | 1·74 | 422 (46) | 10·5 | 169 (45) | 13·3 | | | |
| London I ... ... | 38·2 | 688 (51) | 16·9 | 2·25 | 664 (43) | 9·3 | 275 (55) | 13·1 | 589 (61) | 12·6 | 1·84 |
| London III (*a*) ... | 33·3 | 249 (56) | 6·4 | 1·04 | 521 (30) | 7·5 | 184 (63) | 11·4 | 728 (27) | 19·1 | 2·60 |
| (*n*) ... | 38·6 | 472 (125) | 11·2 | 1·63 | 644 (16) | 8·5 | 231 (39) | 11·9 | 660 (130) | 16·5 | 2·30 |
| Lower Lias, Ser. I ... | 40·7 | 316 (17) | 6·6 | 1·15 | 640 (32) | 7·2 | 161 (37) | 10·1 | 479 (64) | 8·3 | 1·43 |
| Ser. II (*a*) | 31·1 | 321 (54) | 9·2 | 1·46 | 635 (47) | 11·8 | 81 (4) | 8·8 | 460 (53) | 18·2 | 2·60 |
| Carse, Ser. I ... ... | 42·8 | 555 (46) | 11·8 | 1·78 | 840 (75) | 10·9 | 208 (56) | 11·1 | | | |
| Ser. II (*a*) ... | 33·0 | 371 (47) | 10·0 | 1·60 | 552 (112) | 7·8 | | | | | |
| Kirton-in-Lindsey (*a*) | 54·5 | 473 (77) | 8·9 | 1·42 | 918 (128) | 8·1 | | | | | |
| (*n*) | 62·7 | 736 (100+) | 11·2 | 1·68 | 1,290 (84+) | 11·9 | | | | | |
| Gloucester (*a*) ... | 40·8 | 417 (55) | 9·0 | 1·46 | 265 (71) | − 2·2 | | | | | |
| (*n*) ... | 58·8 | 603 (88) | 9·1 | 1·46 | 522 (74) | 0·1 | | | | | |
| Gezira ... ... | 65·0 | 859 (34) | 12·1 | 1·73 | 1,477 (11) | 14·3 | | | | | |
| Mean ... ... | — | — | 10·0 | 1·52 | — | 10·2† | — | 11·1 | — | 14·6 | 2·11 |

* The figures in brackets give the variation from the mean obtained by taking the extreme values of each determination. For example, the individual determinations of the specific volume of the calcium and sodium Trias clays in tetralin are

Ca 0·35556, 0·35548, Na 0·35421, 0·35433

giving as the extreme values for Ca–Na 0·00135 and 0·00115. Since the mean value of the difference is 0·00125, this gives, in the above notation, a difference of 125, with an extreme variation of $\pm 10$ from this mean.

† The figures for the Gloucester clays omitted.

‡ Assuming $V_W = 15{\cdot}5$ cm.

clay is, on the above assumptions, about the same as their difference in crystals, this does not hold for the sodium-calcium or magnesium-calcium differences.

A possible explanation of this discrepancy is the neglect of the water molecules attached to the exchangeable ions on the clay, for it has so far been implicitly assumed in the calculations that the exchangeable ions on the dried clays used were either themselves dry or all contained the same number of molecules of water of hydration, an assumption that is almost certainly false. The effect of these water molecules on the specific volume of the different ionic clays can easily be evaluated. Let each type of ion X have $\lambda_x$ molecules of water attached to each equivalent of X, and let $V_W$ be the mean volume occupied by 1 mole of water of hydration, and further assume that $V_W$ is independent of the exchangeable ion, then using the same notation as before, the specific volumes of a clay holding B equivalents of ions X and Y respectively are

$$\sigma_x = \frac{V + B(V_x + \lambda_x V_W)}{1 + B(M_x + 18\lambda_x)} \quad \text{and} \quad \sigma_y = \frac{V + B(V_y + \lambda_y V_W)}{1 + B(M_y + 18\lambda_y)}$$

giving as the equivalent of equation (1)

$$(\sigma_y - \sigma_x)/B + \sigma_x(M_y - M_x) = (V_y - V_x) + (\lambda_y - \lambda_x)(V_W - 18\sigma_x). \quad . \quad . \quad . \quad (2)$$

Now the quantity on the left-hand side of this equation has been tabulated in Table I under the heading $(V_y - V_x)$. If now this quantity is written as $\Delta V$ (obs) and the true value of $(V_y - V_x)$ is written as $\Delta V$ (true), then

$$\lambda_y - \lambda_x + \frac{\Delta V \text{ (obs)} - \Delta V \text{ (true)}}{V_W + 18\sigma_x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If it is assumed that $\Delta V$ (true) is the value of $(V_y - V_x)$ calculated from crystal density data, and if $V_W$, the volume of 1 mole of water of hydration, is assumed to have the same value as 1 mole of water of crystallization has in a crystal, then $\lambda_y - \lambda_x$, the difference in hydration between one equivalent of Y and of X can be calculated from this equation. Unfortunately $V_W$ is not easy to determine, as the addition of water of crystallization to a crystal usually alters the system to which it belongs. Ignoring this alteration of system, it seems to vary from 13-18 in most substances examined, though usually it lay between 14-16 ccm with an average of about 15·5 ccm. The values of $\lambda_{Ca} - \lambda_{Na}$ and of $\lambda_{Mg} - \lambda_{Ca}$ have been calculated for all the clays available by using the crystal values $-3{\cdot}3$ ccm and $-4{\cdot}0$ ccm for $\Delta V$ (true) and by using $V_W = 15{\cdot}5$ ccm, and are given in Table I. As is seen from this table, the mean values of $\lambda_{Ca} - \lambda_{Na}$ and of $\lambda_{Mg} - \lambda_{Ca}$ are 1·52 and 2·11 respectively. Thus, one equivalent of calcium has, on the average, 1·5 more moles of water than one equivalent of sodium, so that 3 more molecules of water of hydration are attached to each calcium ion than to each sodium ion. Similarly, each magnesium ion has about 4 more molecules of water of hydration than each calcium ion. As a consequence of the assumption that the specific volume

of a clay measured in tetralin is its true specific volume, then on a carefully dried clay the sodium and potassium ions have the same number of molecules of water of hydration, the calcium ion has about 3 more, and the magnesium ion $3 + 4 = 7$ more molecules.

An attempt was made to obtain independent evidence for the correctness of this conclusion by igniting the dry Trias and London I clays in an electric furnace at 650° C for 2 hours, in the hope that at this temperature the ions would have lost all their water of hydration, though unfortunately, owing to lack of time, the clay samples could only be dried at 110° C for 20 hours instead of being much more intensively dried over fresh $P_2O_5$ for several weeks in a vacuum desiccator. The results for the two clays were consistent, and showed that the magnesium ion has 3·5 more molecules of water of hydration than the potassium ion, the calcium ion 2·2 more, and the sodium only 0·5 more. Quantitative agreement between the results of this method and those derived from the specific volume of the clays in tetralin would not be expected, but both methods show that the exchangeable ions on an intensively dried clay can hold water of hydration.

*The specific volume of clays in other non-polar liquids*—Only a few experiments have been made on the apparent specific volume of clays in other non-polar liquids besides tetralin, but the results indicate a sharp difference in behaviour between soil and clay crumbs. Tschapek (*loc. cit.*) using two soils and four non-polar liquids, and the author using only one soil (Rothamsted) and six liquids non-polar or of very slight polarity (benzene, toluene, xylene, paraffin, tetralin and decalin) found that the specific volume of the soil was the same in all the hydrocarbons used within the limits of experimental error. For the few clays examined, which include a Rothamsted separated clay, the apparent specific volumes of dried clay crumbs are different in different hydrocarbons, being largest in decalin and carbon tetrachloride, next in tetralin and paraffin, then in xylene, and smallest in benzene. These differences, though definite, are not large, as all these specific volumes are larger than the specific volume of the dry crumbs in nitrobenzene, which gives the largest specific volume of the polar liquids used. The simplest explanation of this result is that the exchangeable ions do not interact with the hydrocarbons, but that appreciable capillary condensation of the liquid occurs in the pores of the dry clay crumbs though not in the pores of the dry soil crumbs. This explanation needs the assumption that the pores in a soil crumb are larger than in a clay crumb, an assumption that is almost certainly true, although not enough is known about the actual pore size in the crumbs to judge if this is sufficient to cause the difference.

*Specific volume of clays in polar liquids*—This discussion will be limited to the differences between the specific volumes of a clay in water, amyl alcohol, aniline or nitrobenzene and in tetralin. Let T and $L_{1\ 2\ 3\ or\ 4}$ be the specific volumes of a clay in tetralin and in one of the four liquids just mentioned, then (T — L) will give a quantity from which the density of the clay substrate and equivalent weights and volumes of the

exchangeable ions have been eliminated, and which is therefore controlled by differential capillary condensation effects and by the differential interaction between the clay and the two liquids. These quantities (T — L) have the following properties :—

(1) for a given exchangeable ion on the clay (T — $L_1$) is proportional to (T — $L_2$) for different clays. Fig. 3 shows that for hydrogen clays (T — N) is approximately

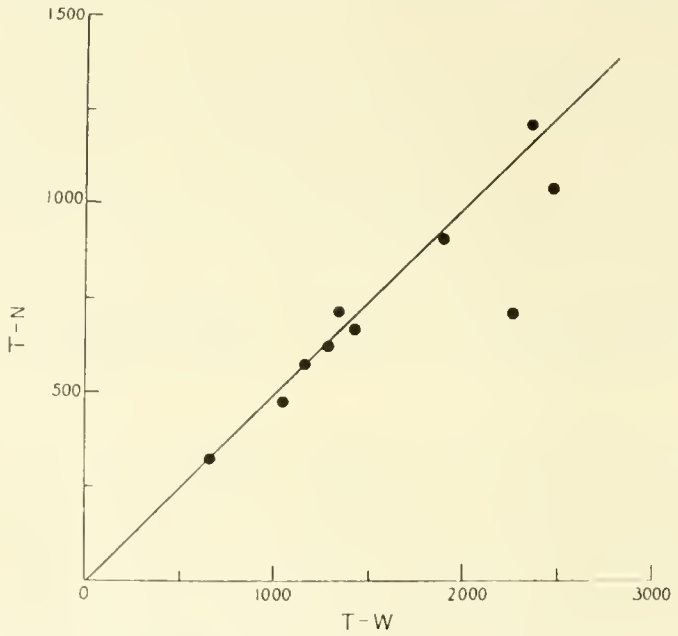



Fig. 3.—Specific volume reduction of hydrogen clays in nitrobenzene compared to water.

proportional to (T — W), where T, N, W are the specific volumes of the clay in tetralin, nitrobenzene, and water respectively ; and fig. 4 shows that for both hydrogen and calcium clays (T — Am) is approximately proportional to (T — W) where Am is the specific volume of the clay in amyl alcohol ;

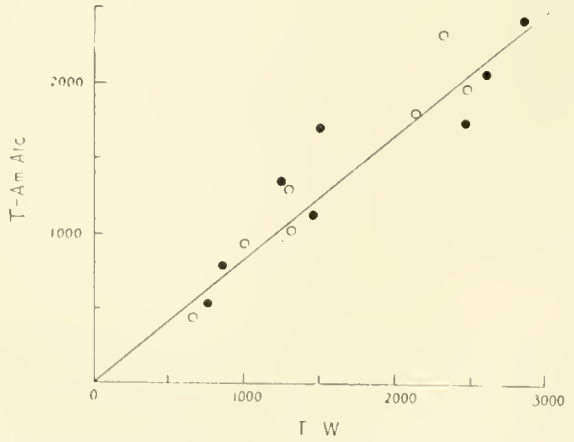



Fig. 4.—Specific volume reduction of hydrogen and calcium clays in amyl alcohol compared to water.
o hydrogen clays ; ● calcium clays.

(2) for a given liquid and a series of clays (T — L) for one exchangeable ion is proportional to (T — L) for a second. This is illustrated in figs. 5 and 6, in which (T — W) for the calcium clays is plotted against (T — W) for the corresponding sodium, potassium, magnesium and hydrogen clays, and in figs. 7, 8, 9, in which (T — L) for the calcium clay is plotted against (T — L) for the corresponding hydrogen clays where L is nitrobenzene, aniline, or amyl alcohol. These two generalizations, though only approximately true, suggest that this quantity (T — L) is largely controlled by the exchangeable ions on the clay, and not by the clay substrate itself.

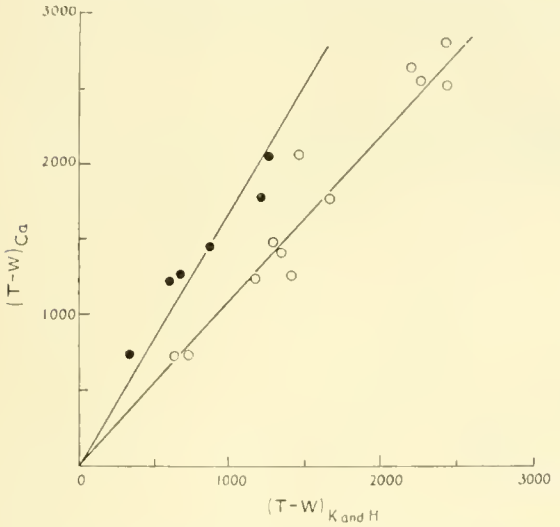



Fig. 5.—Specific volume reduction of potassium and hydrogen clays in water compared to calcium clays. ● Potassium clays; o hydrogen clays.

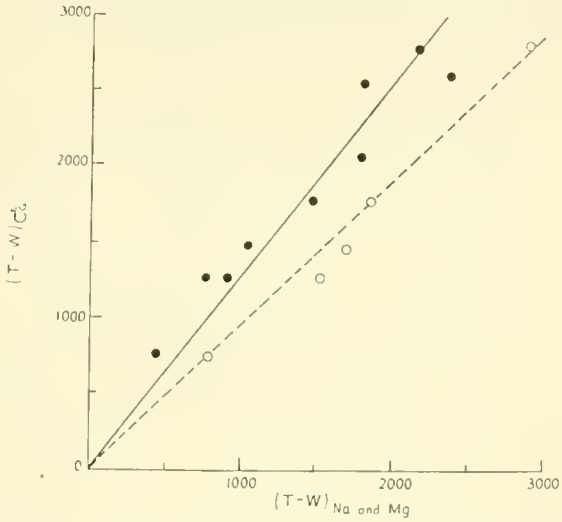



Fig. 6.—Specific volume reduction of sodium and magnesium clays in water compared to calcium clays. ● Sodium clays; o magnesium clays.

(3) (T — L) depends on the base holding power of the clay, since (T — W) for the hydrogen clay is nearly proportional to its base holding power at $p_H$ 7, fig. 10, and since (T — W) for those calcium clays which are nearly saturated up to $p_H$ 7 with calcium is nearly proportional to the calcium held by the clay, fig. 11. Thus (T — L) is approximately proportional to the exchangeable ions held up to $p_H$ 7. The ratio of these two quantities, which gives the mean reduction of specific volume of the clay in water from that in tetralin per equivalent of exchangeable ion, will be called the mean interaction value of that ion.

To investigate what proportion of the total hydrogen ions, that can be exchanged from a clay, take part in this interaction, the relative constancies of the different ratios (T — W)/B ($x$) for all the hydrogen clays available have been compared, where B ($x$) is the equivalents of calcium one gram of hydrogen clay can take up from a calcium buffer solution of $p_H$ ($x$) and N/50 in calcium. The following variates were used for B ($x$) :—*

* For a description of the experimental methods used to measure the various B ($x$), see Schofield, R. K., 'J. Agric. Sci.,' vol. 23, 252 (1933) and 'Trans. VIth Comm. Int. Soc. Soil Sci.,' B, p. 80 (1933).

B (HCl — Ac), B (HCl — N) and B (HCl — P). These variates denote the additional equivalents of calcium held per gram of clay in a solution of half-neutralized calcium acetate ($p_H$ 4·5), half-neutralized calcium *p*-nitrophenol ($p_H$ 7), or half-neutralized calcium phenol ($p_H$ 9·8), respectively, in excess of the bases held by the clay in a solution of 0·04 N HCl and 0·01 N metallic chloride ;

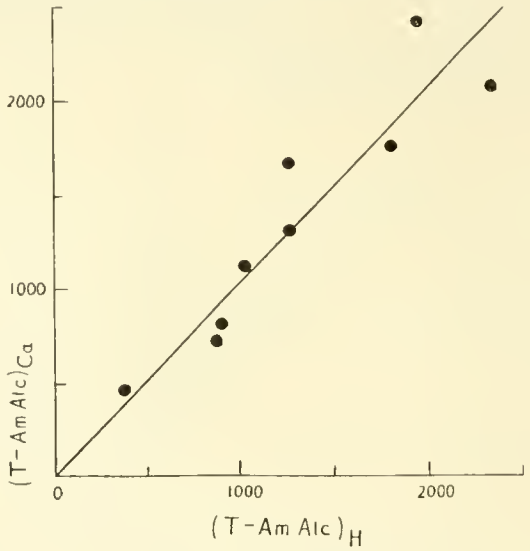



Fig. 7—Amyl alcohol

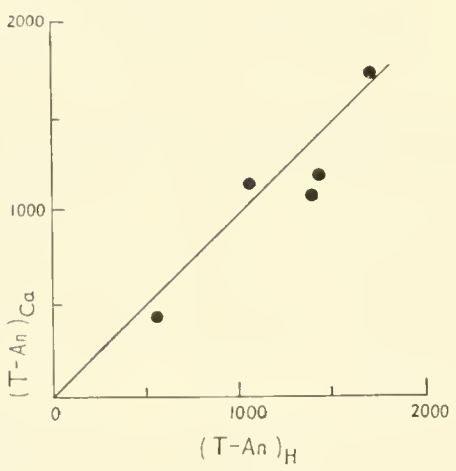



Fig. 8—Aniline

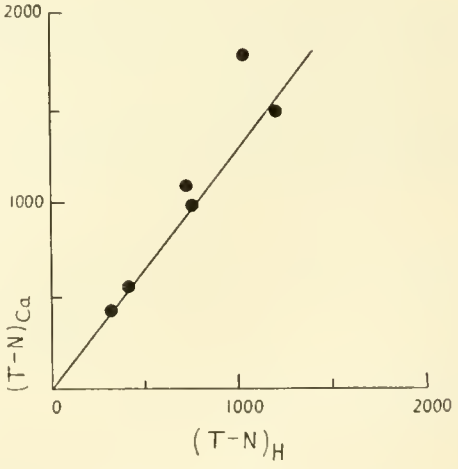



Fig. 9—Nitrobenzene

Figs. 7, 8, 9.—Specific volume reduction of hydrogen clays in different liquids compared to calcium clays.

B (Chl Ac — N), denotes the corresponding additional quantity of calcium held in the nitrophenol solution in excess of the calcium held in a half-neutralized solution of calcium chloracetate ($p_H$ 2·8) ;

B (N), B (P), denote the equivalents of calcium taken up per gram of clay from the

*p*-nitrophenol and the phenol buffers respectively. For every hydrogen clay used these two fall between the hydrochloric acid and the chloracetate values.

In Table II are given the individual values of (T — W)/B ($x$) for all the hydrogen clays available, and the relative constancy of these ratios can be compared by computing the percentage deviation 100 S/M of the individual clays from their mean M for each ratio, where S is the standard deviation of the individual (T — W)/B ($x$) values for the clays from their mean M. The smallest percentage deviation is shown by using B (Cl Ac — N) for B ($x$), while B (HCl — N) and B (N) are almost as good. The three quantities using the phenol buffer point, of which only two are given in the table, give definitely larger deviations, while the acetate buffer point is obviously useless.

The variations of (T — W)/B ($x$) from its mean are probably due to some property of the clay and not solely to chance differences in the amounts of residual bases left in

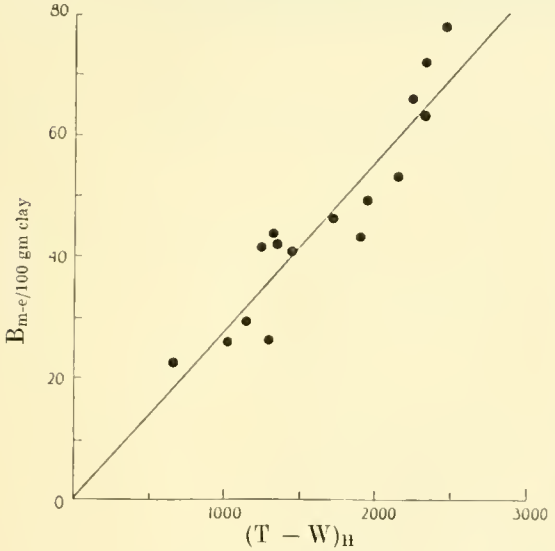



FIG. 10.—Specific volume reduction of hydrogen clays in water to their base holding capacity at $p_H$ 7.

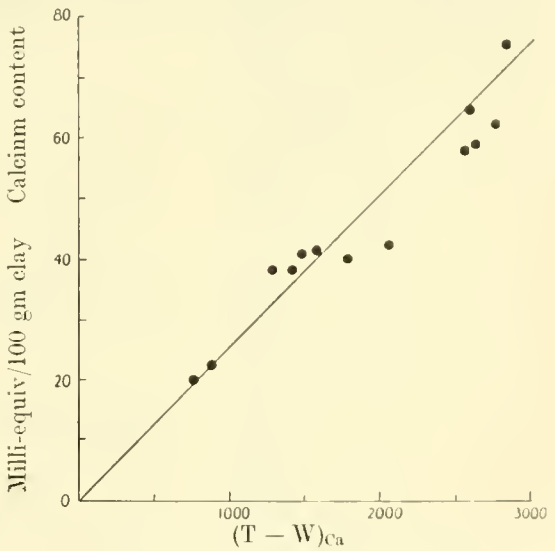



FIG. 11.—Specific volume reduction of calcium clays in water to their calcium content.

the clay and the experimental errors in the determination of (T — W) and the B's, for the values of $\frac{T - W}{B(y - N)}$, where $y$ is either HCl, the chloracetate buffer, or nothing, are strongly correlated with the quantities $\frac{B(HCl—Ph)}{B(HCl—Ac)}$, $\frac{B(HCl—N)}{B(HCl—Ac)}$ and $\frac{B(ClAc—N)}{B(ClAc—Ac)}$. As an example, the scatter diagram of $\frac{T - W}{B(HCl - N)}$ with $\frac{B(HCl - N)}{B(HCl - Ac)}$, which was chosen because it gives the best correlation, is given as fig. 12. In this figure five points have been ringed, namely, two points for the Lower Lias clays, and one for each of the Carse, Kirton-in-Lindsey (dried) and Rothamsted clays. For each of these clays there is a duplicate determination lying nearer the regression line. If these

TABLE II—(T − L)/B ($x$) for the Hydrogen Clays

| | B(HCl—N) × $10^5$ | $\frac{B(HCl-N)}{B(HCl-Ac)}$ | T—W × $10^5$ | (T − W)/B($x$) | | | | | | (T − L)/B (HCl—N) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | B(HCl—Ac) | B(HCl—N) | B(HCl—P) | B(ClAc—N) | B(N) | B(P) | Amyl Alcohol | Aniline | Nitro-benzene |
| Trias (*a*) ... ... ... | 23·8 | 1·63 | 733 | 50·2 | 30·8 | 19·1 | 39·2 | 34·2 | 21·1 | 16·7 | 22·7 | 14·2 |
| (*b*) ... ... ... | 23·8 | 1·63 | 661 | 45·3 | 27·8 | 17·3 | 35·4 | 31·3 | 19·1 | — | — | 13·8 |
| Old Red Sandstone ... | 31·8 | 1·64 | 1,138 | 58·9 | 35·8 | 24·7 | 45·2 | 38·6 | 26·9 | — | — | 17·9 |
| Rothamsted (*a*) ... ... | 39·7 | 1·70 | 1,013 | 43·6 | 25·6 | 15·6 | 36·9 | 37·1 | 17·7 | — | — | 15·6 |
| (*b*) ... ... | 39·7 | 1·70 | 1,191 | 51·1 | 30·0 | 18·4 | 43·3 | 39·9 | 20·8 | — | — | — |
| Lower Lias Ser. I (*a*) ... | 47·5 | 1·62 | 1,679 | 57·5 | 35·4 | 24·9 | 48·0 | 39·2 | 26·5 | — | — | 8·7 |
| (*b*) ... | 47·5 | 1·62 | 1,643 | 56·1 | 34·5 | 24·4 | 46·9 | 38·3 | 26·1 | — | — | — |
| Ser. II ... | 51·9 | 1·80 | 1,702 | 59·3 | 32·9 | 21·9 | 45·9 | 37·5 | 23·9 | — | — | — |
| Carse, Ser. I ... ... | 48·1 | 1·82 | 1,468 | 59·8 | 30·6 | 21·0 | 42·6 | 34·0 | 22·9 | — | — | 19·5 |
| Ser. II ... ... | 60·0 | 1·80 | 2,131 | 64·0 | 35·5 | 24·6 | 48·2 | 40·7 | 26·9 | 29·8 | — | — |
| London I. (*a*) ... ... | 47·0 | 1·60 | 1,415 | 48·1 | 30·1 | 20·9 | 43·2 | 33·8 | 22·6 | 21·3 | 29·6 | 14·3 |
| (*b*) ... ... | 47·0 | 1·60 | 1,325 | 45·1 | 28·2 | 19·6 | 40·4 | 31·6 | 21·2 | — | — | 15·2 |
| London III ... ... | 49·2 | 1·35 | 1,306 | 35·9 | 26·5 | 16·9 | 43·7 | 30·4 | 18·4 | 25·6 | 28·6 | — |
| Gloucester (fine) ... ... | 68·4 | 1·99 | 2,240 | 65·1 | 32·7 | 22·9 | 40·5 | 36·0 | 24·4 | — | — | 10·5 |
| (coarse) ... | 69·3 | 2·06 | 2,306 | 68·4 | 33·3 | 23·0 | 40·8 | 36·8 | 24·6 | — | — | — |
| Kirton-in-Lindsey (dry) ... | 70·6 | 1·65 | 2,451 | 57·2 | 34·6 | 25·8 | 46·7 | 39·0 | 27·7 | — | — | — |
| (freeze) | 70·6 | 1·65 | 2,214 | 51·6 | 31·3 | 23·3 | 42·2 | 33·9 | 25·0 | — | 23·6 | — |
| Gezira ... ... ... | 84·6 | 1·38 | 2,328 | 38·9 | 27·5 | 19·5 | 39·2 | 36·2 | 23·5 | 27·5 | — | 14·2 |
| Palestine ... ... ... | 91·2 | 1·56 | 2,459 | 42·2 | 27·0 | 20·0 | 35·8 | 31·6 | 22·3 | 21·4 | — | 11·5 |
| Fen ... ... ... ... | 124 | 2·42 | 4,645 | 91·0 | 37·4 | 29·1 | 44·1 | 42·6 | 32·2 | — | — | 10·9 |
| Mean ... ... ... | — | — | — | 54·5 | 31·4 | 21·6 | 42·4 | 36·1 | 23·8 | 23·7 | 26·1 | 13·9 |
| 100S/M ... ... | — | — | — | 22·8% | 11·2% | 15·8% | 9·1% | 9·5% | 13·2% | — | — | — |

five " bad " duplicate determinations are ignored there is only one point, that for the Old Red Sandstone clay, that lies appreciably off the lines. None of the other possible $(T - W)/B\,(x)$ values gave such good correlations with any of the B ratios. Thus $(T - W)/B\,(y - N)$, the mean interaction value of each exchangeable hydrogen ion with water, seems to increase as the B ratios, for example, $\frac{B\,(HCl - N)}{B\,(HCl - Ac)}$, increase.

From the physical-chemical standpoint $\frac{B\,(HCl - N)}{B\,(HCl - Ac)}$, namely, the ratio of the calcium which the hydrogen clay holds in a $p_H$ 7 buffer, to the calcium which it holds in a $p_H$ 4·5 buffer, would be expected to measure the strength of the clay acid, this ratio increasing as the strength of the acid decreases. From this point of view, the mean effectiveness of the hydrogen ions on the clay in reducing the specific volume of the clay in water would increase as the strength of the clay acid decreases, a conclusion not easy to interpret in physico-chemical terms. A second possibility is that as the clay acid becomes weaker the interaction between the molecules of the dispersion

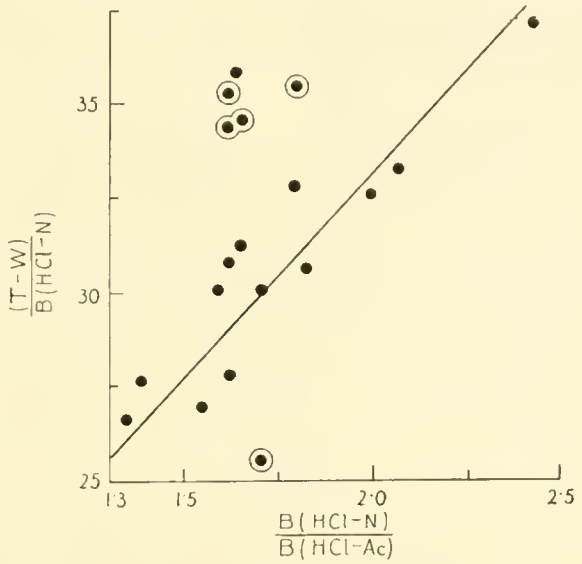



Fig. 12.

liquid and the negative charges on the clay surface becomes stronger. A third possibility is that as the clay acid becomes weaker the pores in the clay crumbs become smaller, so that capillary condensation becomes increasingly more important. This would allow the actual reduction in the specific volume of the clay in water brought about by the hydrogen ions either to be independent of, or to decrease with, the strength of the clay acid, so long as the capillary condensation effect was more or less proportional to $\frac{B\,(HCl - N)}{B\,(HCl - Ac)}$ for the clay. Unfortunately, so far no measurements of the distribution of pore size in clay crumbs have been made.

The effect of other exchangeable ions held by the clay on the reduction of the specific volume of the clay in water cannot be studied so simply as the hydrogen ion effect, as all other ionic clays contain both exchangeable bases and exchangeable hydrogen, so that the effects of the residual hydrogen on the specific volume is always superimposed on the effects of the other exchangeable cations. This complication will be minimized if all the ionic clays used are in equilibrium with a standard buffer solution of as high a $p_H$ as possible. There are, however, experimental difficulties in preparing pure dry clays in equilibrium with buffers above $p_H$ 7 owing to the formation of insoluble calcium and magnesium carbonates, so that the half-neutralized *p*-nitrophenol buffer of $p_H$ 7 was used as standard for preparing the clays. Further, from the preceding discussion on the hydrogen clays, it appeared that this was about as high as it is necessary to go to eliminate the effect of the residual hydrogen. Although at one stage of their preparation, the clays were in equilibrium with this buffer solution, a certain loss of exchangeable base occurs on removal of this solution, so that the dried clay crumbs do not actually contain sufficient base to be in equilibrium with it, but this loss does not usually exceed 3 milli-quivalents per 100 grams of clay. For these reasons, the influence of the exchangeable bases on the specific volume of the clay cannot be accurately determined. If, as before (T — W) is the reduction in specific volume of the clay in water from that in tetralin, one can, as a first approximation, assume either that it is due to the exchangeable ions actually present and ignore the residual hydrogen, or that it is due both to the exchangeable ions present and to the residual hydrogen, with the implicit assumption that the latter behaves like the base. On the first assumption the mean effect of one equivalent of base held by the clay is (T—W)/B, where B is the amount of base actually held, and on the second is (T — W)/B (HCl — N). Actually the difference between these two quantities is small, and the results are qualitatively the same on either assumption, so that only the results based on the second assumption will be given.

Table III gives the values of (T — W)/B (HCl — N) for all the available non-hydrogen clays which contained nearly sufficient base to be in equilibrium with a $p_H$ 7 buffer, and Tables II and III also give the values of (T — L)/B (HCl — N) for the hydrogen and calcium clays when L is the specific volume of the clay in amyl alcohol, aniline or nitrobenzene. As with the hydrogen clay, (T — L)/B (HCl — N) is fairly constant. The individual variations from the mean can be seen, on plotting out, to be dependent on the clay, and thus on $\frac{\text{B (HCl — N)}}{\text{B (HCl — Ac)}}$. Naturally, in considering a plot of two quantities such as (T — L)/B (HCl — N) against one another, experimental errors are very serious, for not only are the (T — L)'s and the individual B's subject to appreciable errors, but, since the degree of saturation of the clays with base is variable, this also produces a source of error in the clay.

A certain number of experiments were made with clays holding less base than was necessary to bring them into equilibrium with a $p_H$ 7 buffer. The clays were intended

to be in equilibrium with half-neutralized acetate buffer of $p_H$ 4·5 containing N/50 base and N/25 acid, but they were found to have taken up too much base, the extra proportion taken up varying from clay to clay. The results of the comparisons are that for the London, Gloucester, and Lower Lias clays (T — L) seems to be the same for the clay only partially and almost completely saturated up to $p_H$ 7, while for the Kirton-in-Lindsey clay (T — L) increases appreciably between these two levels. The specific volume of the partially saturated clay in tetralin always lies between that of the hydrogen and the fully saturated clay. This indicates that not all the exchangeable ions are equally effective in causing this reduction of specific volume, but that those which can be held in a fairly acid buffer have usually a greater effect.

TABLE III—(T — L)/B for the Non-Hydrogen Clays

| | B(HCl — N) × $10_5$ | (T — W)/B (HCl — N) | | | | (T — L) B/(HCl — N) for Calcium | | |
|---|---|---|---|---|---|---|---|---|
| | | Calcium | Magnesium | Sodium | Potassium | Amyl Alcohol | Aniline | Nitrobenzene |
| Trias (*a*) ... ... ... | 23·8 | 31·3 | 33·6 | 19·0 | 15·3 | 19·5 | 19·5 | 17·7 |
| (*b*) ... ... ... | 23·8 | 31·4 | — | — | — | — | — | — |
| Rothamsted ... ... ... | 39·7 | 31·4 | — | 22·6 | 16·0 | — | — | — |
| London I (*a*) ... ... ... | 47·0 | 27·2 | 32·0 | 16·8 | 14·8 | 23·2 | 23·0 | 16·0 |
| (*b*) ... ... ... | 47·0 | 30·1 | — | — | — | — | — | 20·6 |
| London III ... ... ... | 49·2 | 30·1 | 33·8 | 20·9 | 17·6 | 33·7 | 23·0 | — |
| Lower Lias Ser. I. ... ... | 47·5 | 35·6 | 36·8 | 29·0 | 24·2 | — | — | — |
| Ser. II. ... ... | 51·9 | 30·0 | — | — | — | — | — | 13·7 |
| Carse ... ... ... ... | 48·1 | 43·0 | — | 37·4 | 26·7 | — | — | — |
| Gloucester Ser. I. ... ... | 64·4 | — | — | — | — | — | — | 15·6 |
| Ser. II. fine ... | 68·4 | 38·4 | — | 35·1 | — | — | — | 15·2 |
| Ser. II. coarse ... | 69·3 | 37·0 | — | 26·0 | — | — | — | — |
| Kirton-in-Lindsey ... ... | 70·6 | 39·1 | — | 30·4 | — | — | 23·9 | — |
| Palestine ... ... ... | 91·2 | 30·9 | 31·8 | — | — | 26·3 | — | 19·1 |
| Fen ... ... ... ... | 124 | 40·1 | — | — | — | — | — | — |
| Mean ... ... ... | — | 34·0 | 33·6 | 26·4 | 19·1 | 25·7 | 22·3 | 16·9 |

The mean values of the quantities (T — L)/B(HCl — N) for all the hydrogen clays and all the basic clays nearly in equilibrium with the $p_H$ 7 buffer are given in Table IV. The values are not all directly comparable, however, since the means are derived by averaging over different clays. For example, if the values for the calcium and magnesium are compared for the same clays in water the values are Ca = 31·0 and Mg = 33·6, showing that magnesium gives a definitely larger specific volume reduction in water than calcium. The general order of ion effectiveness is independent of the wetting liquid, and the order of increasing specific volume reduction brought about by one equivalent of exchangeable ion is

K, Na, Ba, H, Ca, Mg.

TABLE IV—Mean values of (T — L)/B (HCl — N) for different exchangeable ions and different liquids

| | Potassium | Sodium | Hydrogen | Calcium | Magnesium |
|---|---|---|---|---|---|
| Water ... ... | 19·1 (6) | 26·4 (9) | 31·4 (20) | 34·0 (14) | 33·6 (5) |
| Amyl Alcohol ... | 11·3 (1) | 11·4 (1) | 23·7 (6) | 25·7 (4) | 36·1 (2) |
| Aniline ... ... | 9·5 (1) | 10·4 (1) | 26·1 (4) | 22·3 (4) | 33·4 (1) |
| Nitrobenzene ... | 7·3 (1) | 8·5 (2) | 13·9 (12) | 16·9 (7) | 20·0 (2) |

The figures in brackets give the number of clays used to determine the mean value.

Relative to Ca = 1 for each liquid*

| | Potassium | Sodium | Hydrogen | Calcium | Magnesium |
|---|---|---|---|---|---|
| Water ... ... | 0·59 | 0·76 | 0·91 | 1 | 1·08 |
| Amyl Alcohol ... | 0·33 | 0·35 | 0·86 | 1 | 1·21 |
| Aniline ... ... | 0·41 | 0·45 | 1·17 | 1 | 1·46 |
| Nitrobenzene ... | 0·46 | 0·49 | 0·84 | 1 | 1·14 |
| Surface density of charge ... ... | 0·38 | 0·55 | — | 1 | 1·48 |

* In this table only results based on direct comparisons with the same clays have been used, so this table cannot be derived from the upper half.

With the exception of hydrogen there is not a single experiment in all those performed in which this order comes out significantly different. The position of hydrogen clay in water is erratic, for although on the average its effect is less than the calcium, in two experiments it comes above calcium and in two below sodium. The position of barium is based on a comparison of only one clay saturated with different exchangeable ions. This order for the activeness of the exchangeable ions in interacting with water has also been found by PATE*, BAVER,† and ANDERSON‡ who measured the heat of wetting of soils and clays saturated with different exchangeable ions, and by JANERT§ who measured the heat of wetting of two of the separated clays here described (Trias, Lower Lias), so that magnesium clays do seem to interact with water or other polar dispersion medium more actively, and potassium clays less actively, than any of the other ionic clays used.

If the mean values of (T — L)/B (HCl — N) for the different exchangeable ions are calculated relative to the calcium ion for the four liquids given in Table IV, these relative values are more or less independent of the liquid, as is seen in the lower half of the table ; they are primarily a property of the ion and not of the ion-liquid system. These relative values could probably be predicted from the hypothesis that (T — L)/B (HCl — N), the effectiveness of the ion in reducing the specific volume of a clay in a polar liquid, is proportional to the mean surface density of electric charge on the ion, namely, $ze/4\pi r^2$, where

* 'Soil Sci.,' vol. 20, p. 329 (1925).

† 'J. Amer. Soc. Agron.,' vol. 20, p. 921 (1928).

‡ 'J. Agric. Res.,' vol. 38, p. 565 (1929).

§ 'J. Agric. Sci.,' vol. 24, p. 136 (1934).

$r$ and $ze$ are the radius and the electric charge on the ion. This hypothesis cannot be tested directly since the radii of the ions are not known, but if the radii of the metal atoms in the metal crystal, which are known from X-ray data, are used, a test can be made. In the last row of Table IV are given the surface densities of charge on the different ions relative to calcium calculated in this way, the radii of the atoms used in this calculation being Mg 1·61, Ca 1·97, Na 1·86, and K 2·25A.* These figures are seen to be of the same order of magnitude as the effectiveness of the three metallic ions relative to calcium ; thus giving some support to the hypothesis. Exact agreement could not be expected, since the liquids will interact to some extent with the negative charges on the clay surface, and in the above calculation this interaction has all been attributed to the exchangeable ions. No method has yet been used to measure this interaction, which probably varies with the exchangeable ion, for one would expect that the larger the proportion of the exchangeable ions dissociated from the clay surface, the larger will be the number of negative charges on the clay surface that can interact with the molecules of the dispersion medium. Evidence of this effect can be detected in the latter half of Table IV ; for example, the sodium and potassium figures are too high as compared with calcium in water, owing to the larger dissociation of these ions from the clay. There is not yet sufficient data on the relative dissociations of the different ions from clays dispersed in different liquids to make possible a quantitative discussion of these discrepancies.

The conclusion of this part of the experimental results is, therefore, that (T — L) is controlled by three main factors : the liquid L used, the amount and the surface charge density of the exchangeable ions on the clay, and by one subsidiary factor, specifying to some extent the shape of the buffer titration curve of the clay or the strength of the clay acid. This factor is probably connected with the interaction between the negative charges on the clay surface and the molecules of the wetting liquid.

### *The Properties of the Wetting Liquid Affecting the Specific Volume of the Clay*

A few experiments were made with other polar liquids to try and discover what properties of the wetting liquid are connected with this reduction of the clay specific volume. For the liquids considered : water, amyl alcohol, aniline, and nitro-benzene, the order of effectiveness of the liquids in decreasing the specific volume of the clay is nitrobenzene least, then aniline, then amyl alcohol, with water greatest ; and this order is almost independent of the exchangeable ion on the clay, as is seen from Table IV. Hydrogen clay in aniline is the only definite exception, though possibly magnesium clay in amyl alcohol may be one also. Magnesium clay in aniline is not one, since in the only comparison available (T — W) comes out larger than (T — An). The hydrogen exception is definite, as is seen from all the four available comparisons of the hydrogen clays with calcium clays given in Tables II and III, and is due to a specific interaction

* Taylor, "A Treatise on Physical Chemistry," vol. I, p. 163 (1930).

between the hydrogen clay and the amine group of aniline, for the same result is obtained if o-xylidine or dimethyl aniline is used instead. Thus, for the separated London III clay, the reduction of specific volume from the tetralin figure of the hydrogen and calcium clays in these three amines is

| | Aniline | o-Xylidine | Dimethyl aniline |
|---|---|---|---|
| Hydrogen clay ... ... | 0·01410 | 0·01478 | 0·01295 |
| Calcium clay ... ... | 0·01135 | 0·01008 | 0·00857 |
| Difference ... ... | 0·00275 | 0·00470 | 0·00438 |

Presumably a simple chemical action is involved: the amine group opens up and is adsorbed to give a substituted ammonium clay, thus

$$\boxed{\text{Clay}} - \text{H} + \text{NH}_2.\text{C}_6\text{H}_5 \rightarrow \boxed{\text{Clay}} - \text{NH}_3.\text{C}_6\text{H}_5.$$

The only factor affecting the liquid order so far recognized is chemical action of the type just discussed. The liquid order is neither in the order of their dielectric constants nor of the quantities $\mu^2/\varepsilon$ of Ostwald,* where $\mu$ is the dipole and $\varepsilon$ the dielectric constant of the liquid, though with the exception of nitrobenzene it is in the order of their dipole moments which is aniline ($1 \cdot 55 \times 10^{-18}$ e.s.u.), amyl alcohol (1·68), water (? 1·8), nitrobenzene (3·89).† It is not possible at present to account for this discrepancy in the order of nitrobenzene.

Differences in dipole moment between two similar liquids whose molecules are about the same size can nearly quantitatively account for the different reductions of the specific volume of a clay in them below the clay's specific volume in tetralin, for under these conditions the reduction in specific volume is approximately proportional to the dipole moment of the liquid. As an example, the specific volumes of two calcium, two magnesium, a hydrogen, and a sodium London III clay were determined in tetralin, aniline, and dimethyl aniline, and the ratios of the reduction in the specific volume in aniline below that in tetralin to the reduction in dimethyl aniline were:

| Hydrogen clay | Magnesium clays | Calcium clays | Sodium clay |
|---|---|---|---|
| 1·09 | 1·12 | 1·33 | 1·24 |
| | 1·08 | 1·35 | |

while the ratio of the dipole moments of aniline and dimethyl aniline is $\frac{1 \cdot 55 \times 10^{-18}}{1 \cdot 4 \times 10^{-18}}$ $= 1 \cdot 11$, which is in good agreement with the hydrogen and magnesium figures, but not so good for the calcium or sodium, though the sodium figure is subject to a large experimental error. A further comparison was made of the specific volumes of the three calcium clays, Lias, London I, and Gloucester in the liquids nitrobenzene, chlorbenzene,

* Ostwald and Ortloff, 'Kolloidzschr.,' vol. 59, p. 25 (1932).

† The dipole moments have, wherever available, been taken from Sutton, 'Proc. Roy. Soc.,' A, vol. 133, p. 685 (1931).

benzyl chloride, and benzyl alcohol, and the results are given in Table V, where N, ClB, BCl, and B Alc denote the specific volumes of the clays in these four liquids.

TABLE V—Comparison of the Reduction of Specific Volume of a Clay in two Liquids with the Ratio of their Dipole Moments

| | $\frac{T-N}{T-ClB}$ | $\frac{T-BAlc}{T-BCl}$ | $\frac{T-BAlc}{T-N}$ | $\frac{T-BCl}{T-ClB}$ |
|---|---|---|---|---|
| Ca-Lias ... ... ... ... | 3·13 | — | 1·20 | — |
| Ca-London I ... ... ... | 3·02 | — | 1·14 | — |
| Ca-Gloucester ... ... ... | 1·09 | 2·06 | 1·31 | 1·26 |
| Ratio of the dipole moments of the liquids ... ... | $\frac{3·93}{1·56} = 2·52$ | $\frac{1·6}{1·82} = 0·88$ | $\frac{1·6}{3·93} = 0·41$ | $\frac{1·82}{1·56} = 1·16$ |

The determination of the specific volume of the Lias and London clays in benzyl chloride failed owing to a violent reaction which took place between the clays and the benzyl chloride, and which resulted in HCl being given off and a red liquid remaining. It is seen that using the liquids nitrobenzene, chlorbenzene and benzyl chloride, the relative reductions of clay specific volumes in them is of the same order as the ratio of their dipole moments, but the liquid benzyl alcohol lies well outside this group. Thus there appear to exist sets of liquids such that differences between the reduction of the specific volume of a clay in any two liquids in a given set can be accounted for nearly quantitatively by the difference of dipole moment of the molecules of the two liquids. The liquid property that determines to what set a given liquid shall belong is not the molecular size, for benzyl alcohol and benzyl chloride, for example, belong to different sets, but is probably the complex property of the liquid that determines how much molecular association takes place, and how much capillary condensation the liquid will suffer in the clay crumbs.

A few experiments were made to estimate the magnitude of the capillary condensation effects with different liquids by comparing the specific volume of the clay when fully dispersed in the liquid with that of the dried clay crumb wetted with the liquid. If $\sigma_d$ and $\sigma_c$ are the specific volumes of a dispersed clay and of the dry clay crumbs, and if $m_c$ grams of liquid suffer capillary condensation per gram of clay crumb, and if $v_c$ is the mean volume of the condensed liquid and $v_o$ of the free liquid, then

$$\sigma_d - \sigma_c = m_c (v_o - v_c) \quad \ldots \ldots \ldots \ldots \quad (1)$$

provided that (1) the crumbs are completely wetted, (2) the weight of imbibed liquid per gram of clay is the same for both systems, and (3) in the determination of the specific volume the clays have the same weight of residual liquid when weighed dry. The effects of the first two provisos on equation (1) are

(1) suppose a volume $v_a$ cc of air is entrapped in the crumbs per gram of clay, then

$$\sigma_d - \sigma_c = m_c (v_o - v_c) - v_a$$

(2) suppose $m_i$ and $m'_i$ grams of imbibed liquid are held per gram of dispersed and crumbed clay, and let $(v_o - \delta v)$ be the mean specific volume of the liquid represented by the weight $(m_i - m'_i)$, then

$$\sigma_d - \sigma_c = m_c (v_o - v_c) - (m_i - m'_i)\delta v$$

showing that $\sigma_c$ or $\sigma_d$ is only affected by the imbibitional liquid if there is an appreciable change of volume of the liquid on imbibition.

The chief limitation of the experiments performed is the difficulty in determining the weight of clay in the dispersion. The weight of clay crumbs used for a specific volume determination was taken as the weight of the crumbs after they had been drying in a desiccator over phosphorus pentoxide *in vacuo* for about three weeks. The weight of clay used in the water dispersion was determined by removing the main bulk of the water by evaporation on a water bath, drying at 110° C for 24 hours, and then drying over phosphorus pentoxide *in vacuo* for about a fortnight. Under these conditions it is fairly certain that the moisture content of the two samples of clay will agree within a few tenths of a milligram per gram of clay. The effect of this difference can easily be calculated, for if the dry clay from the dispersion had a moisture content $m_r$ per gram of clay more than the dry clay crumbs, then

$$\sigma_d - \sigma_c = m_c (v_o - v_c) + m_r (v_o - \sigma) \text{ (approx.)},$$

where $\sigma$ is the specific volume of the clay. Thus, if $m_r = 10^{-4}$ grams and $\sigma = 0{\cdot}35$, probably a fair average value, then $\sigma_d - \sigma_c = m_c (v_o - v_c) + 3{\cdot}5 \times 10^{-5}$. Hence each tenth of a milligram difference of moisture content per gram of clay between the two dry clays gives a difference of $3{\cdot}5 \times 10^{-5}$ between the two specific volumes.

The results of the comparisons between the specific volumes of deflocculated clays in water and of their crumbs in water are given in Table VI, which shows that $\sigma_c$ is usually greater than $\sigma_d$, so that if much capillary condensation is taking place, this

TABLE VI—Comparison of the Specific Volume of Deflocculated and Crumbled Clay in Water

| | Specific volume of deflocculated clay $\sigma_d$ | Specific volume of crumbed clay $\sigma_c$ | Difference $(\sigma_d - \sigma_c) \times 10^4$ |
|---|---|---|---|
| Kirton-in-Lindsey NaAc ... | 0·3647 | 0·3656 | − 9 |
| NaN ... | 0·3605 | 0·3602 | + 3 |
| Carse Na ... ... ... | 0·3792 | 0·3800 | − 8 |
| London I H ... ... ... | 0·3547 | 0·3570 | −23 |
| Ca ... ... ... | 0·3536 | 0·3547 | −11 |

effect is more than counterbalanced by incomplete wetting of the crumbs, or to a smaller imbibition of water by the crumbled clay than by the dispersed clay, or to incomplete drying of the clay before wetting. Since none of the last three factors can be very large, only a little capillary condensation can be taking place.

Turning to the results of this comparison for non-aqueous dispersions, the technique was very unsatisfactory, since, although the results showed that appreciable capillary condensation appeared to take place if the dried clay crumbs were wetted with ethyl or amyl alcohol, two sources of uncertainty enter rendering the results doubtful. In the first case the density of the free liquid in the clay dispersion always had a higher density than the pure liquid, and in the second the clay crumbs were weighed dry before being wetted with the organic liquid, while the dispersed clay was weighed after it had been dried from the liquid, so that errors of one or two milligrams per gram of clay could easily arise between the comparative weights of the same clay sample, due to different residual liquids on the dry clay. Thus, although capillary condensation probably takes place in clay crumbs in organic liquids, the results give no certain proof that this is so.

The author is indebted to Dr. B. A. Keen and also to Dr. R. K. Schofield for many helpful suggestions during the course of this work.

*Summary*

Clay particles can form strong aggregates, or crumbs, when dry, only if the clay particles are sufficiently small, if there are a sufficient number of small exchangeable ions on the clay, and if the clay has been dried from a dispersion medium whose molecules are polar and sufficiently small.

Individual clay particles have an appreciable interaction with the molecules of polar liquids ; they do not interact, or only slightly with non-polar liquids. On the hypothesis that no differential interaction occurs between the exchangeable ions on the clay and a non-polar hydrocarbon, the magnesium ions on the clay have about seven and the calcium ions about three more molecules of water of hydration than the sodium or potassium ions. Independent evidence is given for this conclusion.

Clays interact with water and organic liquids containing a polar group by causing a contraction in volume of the liquid. For homionic clays, saturated up to $p_H$ 7 with base, this contraction is nearly proportional to the number and mean charge density on the surface of the exchangeable ions on the clay and, to a less extent, it varies with a parameter specifying the shape of the titration curve of the clay.

These results have been interpreted on the hypothesis that cations can orientate polar molecules of the dispersion liquid around them, and this power is proportional to their surface density of charge. This power is also possessed by the free negative charges on the clay particle. When the dispersion liquid has nearly all been removed the cations bind the negative charges on two clay particles together by means of bridges of strongly orientated molecules of the polar dispersion liquid.

Printed in London, Great Britain, by Harrison and Sons, Ltd., St. Martin's Lane, W.C.2.

Author Rothamsted memoirs
313829
Title (from Philosophical Transactions) 4to. ser. A
4(1914-34)
Biological P
& Medical
Serials R
DATE.
NAME OF BORROWER.
University of Toronto
Biological
& Medical
Serials
Library
DO NOT
REMOVE
THE
CARD
FROM
THIS
POCKET
Acme Library Card Pocket
LOWE-MARTIN CO. LIMITED